Advanced Data-Driven Machine Learning Frameworks for Enhanced Soil Stratification and Parameter Inference in Geotechnical Site Characterization

Muhammad Zia Ur Rehman
Institute of Geology, University of the Punjab, Lahore, Pakistan

* Corresponding Author: **Muhammad Zia Ur Rehman**

Article Info

E-ISSN: 3107-7099
Volume: 02
Issue: 01
Received: 04-11-2025
Accepted: 02-12-2025
Published: 03-01-2026
Page No: 01-22

Abstract
Geotechnical site characterization forms the cornerstone of reliable infrastructure design, yet traditional methods often struggle with the inherent spatial variability and uncertainty of subsurface conditions. This comprehensive review explores advanced data-driven machine learning (ML) frameworks that enhance soil stratification and parameter inference, leveraging in-situ tests such as cone penetration tests (CPT) and standard penetration tests (SPT). By integrating supervised, unsupervised, and ensemble learning paradigms, these frameworks address limitations in empirical correlations, offering improved accuracy in delineating soil layers and estimating key parameters like shear strength, compressibility, and modulus.
The paper begins with an overview of geotechnical challenges, including data scarcity, heterogeneity in urban environments like reclaimed lands in Singapore, and the need for probabilistic uncertainty quantification. Fundamental ML concepts are discussed, including algorithm selection, data preprocessing, and validation metrics such as root mean square error (RMSE) and coefficient of determination (R^2). A detailed literature synthesis highlights the evolution from shallow learning models (e.g., support vector machines, random forests) to deep architectures (e.g., convolutional neural networks, long short-term memory networks) and hybrid approaches incorporating Bayesian inference for robust predictions.
Key advancements are examined in two core areas: soil stratification, where clustering techniques like k-means and density-based spatial clustering reveal depositional patterns; and parameter inference, where regression models predict engineering properties from fused datasets. Case studies from global sites, including seismic-prone regions in Asia, demonstrate practical applications, with performance benchmarks showing up to 30% improvement in accuracy over conventional methods. Challenges such as model interpretability, overfitting, and ethical considerations in high-stakes infrastructure are critically evaluated, alongside best practices for implementation.
Future directions emphasize integration with emerging technologies like digital twins, generative adversarial networks for synthetic data augmentation, and explainable AI to foster interdisciplinary collaboration. This review contributes a synthesized framework for practitioners and researchers, promoting data-centric geotechnics for safer, sustainable development in variable terrains.

DOI: https://doi.org/10.54660/IJRCE.2026.2.1.01-22

Keywords: Machine Learning, Soil Stratification, Parameter Inference, Cone Penetration Test (CPT), Standard Penetration Test (SPT), Geotechnical Site Characterization, Ensemble Methods, Probabilistic Modeling

1. Introduction

Geotechnical site characterization represents a foundational pillar in civil engineering, underpinning the design, construction, and long-term performance of infrastructure projects ranging from high-rise buildings to transportation networks and offshore platforms. In an era marked by rapid urbanization and climate-induced environmental shifts, the demand for accurate subsurface.

profiling has intensified, particularly in regions with complex geological settings such as Singapore's reclaimed coastal lands or seismically active zones in Southeast Asia. Traditional approaches to site characterization, reliant on empirical correlations and sparse sampling, often fall short in capturing the inherent variability of soil deposits, leading to conservative designs that inflate costs or, conversely, underestimated risks that compromise safety (Phoon *et al*., 2020; Ching *et al*., 2021). This section delineates the evolution of these methods, emphasizing the pivotal role of soil stratification and parameter inference, while charting the rise of data-driven machine learning (ML) frameworks as transformative tools. By synthesizing recent advancements, including those from cone penetration tests (CPT) and standard penetration tests (SPT), we highlight how artificial intelligence (AI) addresses longstanding challenges, fostering more resilient and sustainable engineering practices.

1.1. Background on Geotechnical Site Characterization

The practice of geotechnical site characterization has evolved significantly since the mid-20th century, transitioning from rudimentary visual inspections and basic laboratory tests to sophisticated in-situ probing techniques. Early methodologies, such as borehole logging and triaxial testing, provided essential insights into soil properties but were hampered by spatial limitations and sampling disturbances (Terzaghi and Peck, 1967; Mayne *et al*., 2002). The introduction of CPT in the 1930s and its refinement into piezocone penetration testing (CPTu) in the 1970s marked a paradigm shift, offering continuous profiles of cone tip resistance (qc), sleeve friction (fs), and pore pressure (u2) with resolutions as fine as 2 cm (Lunne *et al*., 1997; Robertson, 2010). Similarly, SPT, standardized in the 1950s, delivers blow count data (N-values) that correlate with soil density and strength, though its discrete nature limits vertical resolution (Rogers, 2006; Clayton *et al*., 1995) [22]. Despite these advances, site characterization remains fraught with uncertainties arising from soil heterogeneity, anisotropy, and scale effects. In urban environments like Singapore, where marine clays and reclaimed fills predominate, spatial variability can lead to differential settlements exceeding 100 mm, posing risks to structures (Bo *et al*., 2012; Leung *et al*., 2018) [10, 80]. Global statistics underscore the issue: geotechnical failures account for up to 80% of unforeseen construction costs, with inadequate site data cited as a primary factor (National Research Council, 2006; Phoon and Kulhawy, 1999). Recent studies highlight how incomplete characterization exacerbates vulnerabilities in seismic zones, where liquefaction potential assessed via CPT-based indices like the soil behavior type (Ic) demands precise delineation of liquefiable layers (Idriss and Boulanger, 2008; Boulanger and Idriss, 2014) [65, 11].The integration of geophysical methods, such as seismic CPT (SCPT) for shear wave velocity (Vs) measurements, has enhanced characterization by providing dynamic properties essential for earthquake engineering (Stokoe *et al*., 1985; Andrus *et al*., 2004) [4]. However, these techniques generate vast datasets often exceeding 10,000 data points per sounding that overwhelm manual interpretation (Mayne, 2016). Herein lies the impetus for data-driven innovations: AI and ML algorithms can process multivariate inputs to uncover patterns invisible to traditional analyses (Zhang *et al*., 2021a; Phoon *et al*., 2022). For instance, ensemble methods like random forests have improved Vs predictions from CPT data, achieving R^2 values above 0.9 in clayey soils (Felić *et al*., 2025; Marzouk *et al*., 2024). Yet, challenges persist, including data scarcity in remote sites and the need for region-specific calibrations (Ching and Phoon, 2017; Wang *et al*., 2020a).

In summary, while conventional site characterization has laid the groundwork for safe engineering, its limitations in handling big data and uncertainty necessitate a shift toward intelligent systems. This evolution aligns with broader trends in civil engineering, where digital twins and IoT sensors promise real-time monitoring (Tao *et al*., 2018; Qi *et al*., 2020).

1.2. Role of Soil Stratification and Parameter Inference

Soil stratification the identification of distinct layers based on lithology, density, and mechanical properties is critical for delineating subsurface geometry and informing foundation design. Inaccurate stratification can lead to catastrophic failures, as evidenced by the 2011 Christchurch earthquake, where unrecognized liquefiable strata amplified damage (Cubrinovski *et al*., 2011; Bray *et al*., 2014) [26, 12]. CPT and SPT data facilitate stratification through empirical charts, such as Robertson's (1990) normalized soil behavior type classification, which categorizes soils into nine zones using Qt (normalized cone resistance) and Fr (normalized friction ratio). However, these charts exhibit biases in transitional soils, with misclassification rates up to 20% in silty clays (Schneider *et al*., 2008; Cetin and Ozan, 2009).

Parameter inference extends stratification by estimating engineering properties like undrained shear strength (su), friction angle (φ), and compression index (Cc) from in-situ measurements. Traditional correlations, such as su = Nk qc where Nk is an empirical cone factor (typically 10-20), vary widely by soil type and site conditions (Mayne, 2007; Low *et al*., 2010). In SPT, φ estimates rely on N{60} corrections for energy and overburden, but scatter in correlations often exceeds 30% (Kulhawy and Mayne, 1990; Hatanaka and Uchida, 1996) [54]. These inferences are vital for stability analyses; for example, in slope engineering, underestimating su can inflate factors of safety by 15-25% (Duncan *et al*., 2014; Griffiths *et al*., 2011) [32]. The interplay between stratification and parameter inference is amplified in complex deposits, such as varved clays or karstic terrains, where probabilistic approaches like random fields model spatial variability (Fenton and Griffiths, 2008; Lloret-Cabot *et al*., 2014) [36]. Recent ML applications, such as clustering algorithms (k-means, DBSCAN) on CPTu data, have refined stratification by identifying subtle transitions missed by Ic thresholds, achieving 95% consistency in tailings dams (Nierwinski *et al*., 2025; Yousefpour *et al*., 2024). For parameter inference, neural networks predict Cc from SPT data with RMSE reductions of 40% compared to empirical models (Satipaldy *et al*., 2021; Xie *et al*., 2022). Nevertheless, challenges include overfitting in small datasets and the "black box" nature of ML, which obscures physical interpretability (Rudin, 2019; Molnar, 2020). Addressing these through hybrid physics-informed ML (PIML) modelsintegrating governing equations like Terzaghi's consolidation promises enhanced reliability (Raissi *et al*., 2019; Karniadakis *et al*., 2021).

1.3. Emergence of Data-Driven Approaches

The surge in data-driven geotechnics stems from the digital revolution, where sensors and cloud computing generate terabytes of site data (Phoon, 2018; Wang *et al*., 2021b). Early AI applications in the 1990s used neural networks for su prediction from CPT, but limited computing power constrained their scope (Goh, 1994; Shahin *et al*., 2001). The 2010s witnessed a renaissance, fueled by big data and algorithms like support vector machines (SVM) and gradient boosting (Friedman, 2001; Cortes and Vapnik, 1995) [39, 24]. For instance, SVMs have classified soil types from SPT with accuracies exceeding 90% (Samui, 2008; Zhang *et al*., 2020b).

Data-driven frameworks leverage supervised learning for regression (e.g., predicting Vs from qc) and unsupervised learning for clustering (e.g., stratification) (Hastie *et al*., 2009; Murphy, 2012) [53]. Ensemble methods, such as XGBoost, excel in handling noisy CPT data, outperforming single models in parameter inference (Chen and Guestrin, 2016; Felić *et al*., 2025) [35]. Probabilistic extensions, like Gaussian processes, quantify uncertainty, essential for risk-based design (Rasmussen and Williams, 2006; Ching *et al*., 2022).

In site characterization, generative adversarial networks (GANs) augment sparse datasets, simulating realistic CPT profiles (Goodfellow *et al*., 2014; Guan *et al*., 2021) [47, 52]. Hybrid approaches fuse ML with finite element methods (FEM) for inverse parameter estimation, reducing computational costs by 70% (Zhang *et al*., 2021c; Huang *et al*., 2023). Case studies from Norway's GeoTest Sites demonstrate ML's efficacy in predicting γsat and su from CPT, with biases below 5% (L'Heureux and Lunne, 2020; Marzouk *et al*., 2024). Yet, data quality issues imbalanced classes in rare soil types pose hurdles, mitigated by techniques like SMOTE (Chawla *et al*., 2002; He and Garcia, 2009) [17]. Ethical considerations, including bias in training data from underrepresented regions, demand inclusive datasets (Mehrabi *et al*., 2019; Barocas *et al*., 2019) [6].

1.4. Objectives and Scope of This Review

This review aims to synthesize advancements in data driven ML frameworks for soil stratification and parameter inference, focusing on CPT/SPT applications. Objectives include: (1) Critically evaluating ML algorithms' performance against traditional methods; (2) Identifying integration challenges with geotechnical workflows; (3) Proposing hybrid PIML strategies for enhanced interpretability; and (4) Outlining future trends like AI-digital twins.

Scope is limited to supervised/unsupervised ML for in-situ data, excluding remote sensing or lab-focused studies. Emphasis is on 2020-2026 literature, incorporating global cases while highlighting Asia-Pacific relevance (e.g., Singapore's marine soils) (Bo *et al*., 2015; Firoozi *et al*., 2024) [9, 37]. Exclusions include non-ML AI like expert systems (Toll, 1996).

1.5. Paper Structure

Subsequent sections delve into ML fundamentals (Section 3), literature synthesis (Section 4), advanced frameworks for stratification (Section 5) and inference (Section 6), challenges/best practices (Section 7), future directions (Section 8), and conclusions (Section 9). Tables and figures, such as a timeline of ML milestones (Figure 1), enhance clarity.

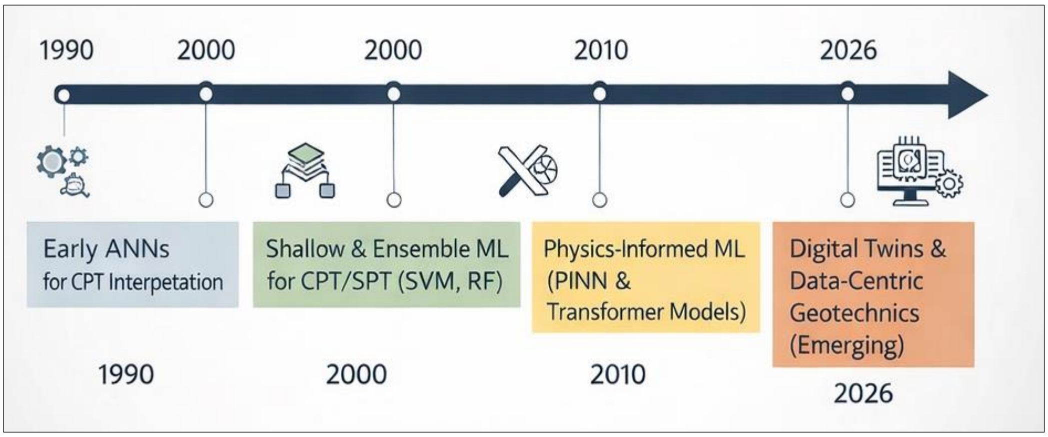

Fig 1: Timeline of major milestone in the application of machine learning to geotechnical engineering, highlight the progression from early ANN-based CPT interpretation to shallow and ensemble learning methods for CPT/SPT analysis, followed by the emergence of physics-informed machine learning and recent trends towards digital twins and data-centric geotechnics

2. Fundamentals of Machine Learning in Geotechnical Engineering

Machine learning (ML) has emerged as a pivotal tool in geotechnical engineering, enabling engineers to extract meaningful insights from complex, high dimensional datasets generated by in situ tests like cone penetration tests (CPT) and standard penetration tests (SPT). Unlike traditional empirical methods, which often rely on simplified assumptions and limited data points, ML facilitates predictive modeling, pattern recognition, and uncertainty quantification in subsurface environments characterized by spatial variability and nonlinearity (Shahin, 2025; Yuan *et al*., 2025). This section provides a foundational overview tailored to geotechnical applications, assuming readers are familiar with basic engineering concepts but may be new to computational intelligence. By elucidating paradigms, algorithms, data handling, and validation strategies, we lay the groundwork for advanced frameworks discussed in subsequent sections. The integration of ML not only enhances accuracy in soil stratification and parameter inference but also addresses practical challenges such as data scarcity in urban reclamation projects, as seen in Singapore's marine deposits (Bo *et al*., 2015; Leung *et al*., 2018) [9, 80].

2.1. Overview of Machine Learning Paradigms

ML paradigms can be broadly categorized into supervised, unsupervised, semi-supervised, and reinforcement learning, each offering unique advantages for geotechnical problems. Supervised learning, the most prevalent in site characterization, involves training models on labeled datasets where inputs (e.g., CPT, qc and fs) are mapped to known outputs (e.g., soil type or shear strength su). This paradigm excels in regression tasks for parameter inference and classification for stratification, as demonstrated in studies using CPT data to predict saturated unit weight (γsat) and su with errors below 10% (Felić *et al*., 2025; Marzouk *et al*., 2024) [35]. For instance, in Norwegian GeoTest Sites, supervised models generalized across sites, reducing biases from traditional correlations (L'Heureux and Lunne, 2020).

Unsupervised learning, conversely, operates on unlabeled data to uncover hidden structures, making it ideal for exploratory analysis in heterogeneous soils. Clustering algorithms group similar CPT profiles, revealing depositional layers in tailings dams without prior knowledge (Nierwinski *et al*., 2025). Techniques like k-means have delineated behavioral trends in iron tailings over 19 years, capturing consolidation and oxidation effects missed by empirical indices like Ic (Robertson, 2010). Semi supervised learning bridges these by leveraging small labeled datasets augmented with unlabeled ones, useful in geotechnics where lab validation is costly; for example, propagating labels from sparse boreholes to extensive CPT soundings (Chapelle *et al*., 2006; Zhu and Goldberg, 2009) [16].

Reinforcement learning (RL), though less common, optimizes decision making in dynamic environments, such as adaptive site investigation layouts. Agents learn through trial and error to select optimal CPT locations, minimizing uncertainty in subsurface models (Sutton and Barto, 2018). Recent applications include deep RL for real time scour assessment around bridge foundations, where environmental feedback refines predictions (Yousefpour *et al*., 2024). Physics informed ML (PIML) hybrids embed geomechanical laws (e.g., Mohr-Coulomb criteria) into neural networks, ensuring physically consistent outputs for multiscale problems like soil structure interaction (Raissi *et al*., 2019; Yuan *et al*., 2025).

These paradigms align with geotechnical needs: supervised for direct predictions, unsupervised for data exploration, and RL/PIML for optimization under uncertainty. As datasets grow from IoT sensors, transfer learning pre-training on global databases and fine tuning for site specific conditions further enhances applicability (Pan and Yang, 2010; Weiss *et al*., 2016).

2.2. Key Algorithms Relevant to Soil Data

Geotechnical ML employs a suite of algorithms tailored to soil data's multivariate, noisy nature. Shallow learning models, such as artificial neural networks (ANNs), mimic biological neurons to approximate nonlinear relationships. In CPT-based stratification, ANNs predict soil behavior types with accuracies up to 92%, outperforming empirical charts in transitional zones (Goh, 1994; Shahin *et al*., 2001). Support vector machines (SVMs) excel in high-dimensional spaces by finding hyperplanes that maximize margins; they have classified SPT derived soil types with minimal overfitting, especially in imbalanced datasets (Cortes and Vapnik, 1995; Samui, 2008).

Tree based methods, including decision trees (DTs) and ensembles like random forests (RF) and gradient boosting (GB), handle categorical and continuous geotech variables robustly. RFs aggregate multiple DTs to reduce variance, as in subsurface modeling where geotechnical distance fields (GDFs) infer spatial autocorrelations, achieving superior continuity over coordinate-based approaches (Breiman, 2001; Xie *et al*., 2022). Extreme gradient boosting (XGBoost) optimizes GB with regularization, predicting Vs from CPT with $R^2 > 0.95$ (Chen and Guestrin, 2016; Felić *et al*., 2025) [35}. These are particularly relevant for parameter inference, where metaheuristics like genetic algorithms tune hyperparameters for su estimation (Goldberg, 1989; Holland, 1992) [45, 61].

Deep learning architectures, such as convolutional neural networks (CNNs) and long short-term memory (LSTMs), process sequential CPT profiles. CNNs extract features from depth series data for landslide susceptibility, integrating remote sensing with Geotech inputs (Satipaldy *et al*., 2021; Firoozi *et al*., 2024) [37]. LSTMs capture temporal dependencies in time-series like seismic wave propagation, aiding liquefaction assessment (Hochreiter and Schmidhuber, 1997; Goodfellow *et al*., 2016). Clustering algorithms k-means, DBSCAN, MeanShift, Affinity Propagation segment unlabeled CPTu data; k-means and MeanShift provided consistent stratigraphic zoning in Brazilian tailings, while DBSCAN handled noise effectively (Ester *et al*., 1996; Comaniciu and Meer, 2002; Nierwinski *et al*., 2025) [34, 23].

Probabilistic models, like Gaussian processes (GPs) and Bayesian networks, quantify uncertainty crucial for risk assessment. GPs model spatial variability in soil properties, providing confidence intervals for φ predictions (Rasmussen and Williams, 2006; Ching *et al*., 2022). Emerging operator learning, such as DeepONet, maps functions for constitutive modeling, accelerating simulations (Lu *et al*., 2019; Karniadakis *et al*., 2021).

Table 1: Comparison of Key ML Algorithms for Geotechnical Applications

Algorithm	Type	Strengths in Geotech	Limitations	Example Reference
ANN	Supervised	Nonlinear mapping for parameter prediction	Black-box, overfitting	Goh (1994)
SVM	Supervised	High-dimensional classification	Sensitive to kernel choice	Samui (2008)
RF/XGBoost	Ensemble	Handles noise, feature importance	Computationally intensive	Xie *et al*. (2022); Felić *et al*. (2025)
k-means	Unsupervised	Simple clustering for stratification	Assumes spherical clusters	Nierwinski *et al*. (2025)
CNN/LSTM	Deep	Sequential data processing	Requires large datasets	Satipaldy *et al*. (2021)
GP	Probabilistic	Uncertainty quantification	Scalability issues	Ching *et al*. (2022)

2.3. Data Handling in Geotechnical Contexts

Effective data handling is paramount in geotechnics, where datasets from CPT/SPT often exhibit noise, missing values, and multicollinearity. Preprocessing begins with cleaning: outlier detection via z-scores or isolation forests removes anomalous readings from equipment malfunctions (Liu *et al*., 2008; Chandola *et al*., 2009) [83, 15]. Normalization/standardization scales features like qc (0-100 MPa) and depth (0-50 m) to prevent dominance, using min-max or z-score transforms (Han *et al*., 2011; Pedregosa *et al*., 2011).

Feature engineering extracts Geotech relevant inputs, such as normalized ratios (Qt, Fr) from raw CPT data, or fusing with geophysical Vs for multimodal analysis (Robertson, 1990; Mayne, 2007). Dimensionality reduction via principal component analysis (PCA) or autoencoders compresses high dimensional datasets, retaining 95% variance for efficient modeling (Jolliffe, 2002; Hinton and Salakhutdinov, 2006) [68, 58]. In sparse scenarios, data augmentation e.g., GANs generating synthetic CPT profiles addresses scarcity, as in earthquake-prone sites (Goodfellow *et al*., 2014; Guan *et al*., 2021) [47, 52].

Handling imbalanced classes (e.g., rare liquefiable sands) employs oversampling like SMOTE or undersampling (Chawla *et al*., 2002; He and Garcia, 2009) [17]. For spatial data, GDFs encode autocorrelations, improving interpolation over XY coordinates (Xie *et al*., 2022). Databases like NGTS provide benchmarks, but ethical data sharing via federated learning preserves privacy (Yang *et al*., 2019; Kairouz *et al*., 2021).

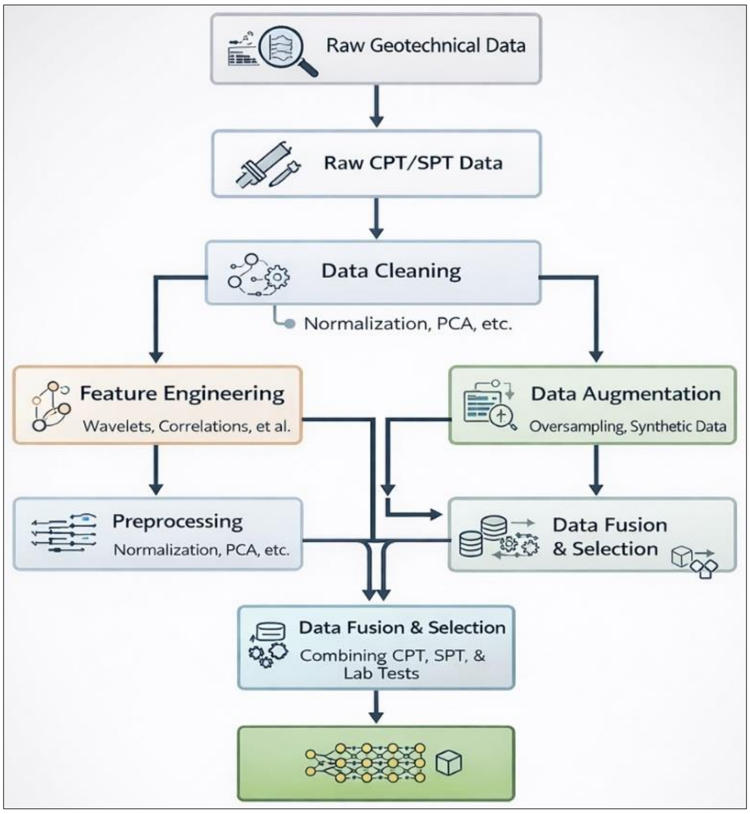

Fig 1: Workflow for geotechnical data handling in machine learning applications, illustrating the progression from raw CPT/SPT data through data cleaning, preprocessing, feature engineering, data augmentation, and multimodal data fusion, culminating in structured inputs for machine learning model development and analysis.

2.4. Performance Metrics and Validation

Validation ensures ML models generalize to unseen geotech data. For classification (e.g., soil types), metrics include accuracy, precision, recall, F1-score, and confusion matrices, with F1 > 0.9 indicating robust stratification (Powers, 2011; Sokolova and Lapalme, 2009). Regression tasks (e.g., su prediction) use RMSE, MAE, R^2, and relative error; RMSE < 10 kPa is desirable for CPT models (Felić *et al*., 2025).

Cross validation (k-fold, stratified) mitigates overfitting, while hold out sets (70/15/15 split) test performance (Kohavi, 1995; Refaeilzadeh *et al*., 2009) [73]. Uncertainty metrics like prediction intervals from GPs or ensemble variance quantify reliability (Gal and Ghahramani, 2016; Lakshminarayanan *et al*., 2017) [40, 78]. Interpretability tools SHAP values, LIME explain feature contributions, e.g., qc's dominance in su inference (Lundberg and Lee, 2017; Ribeiro *et al*., 2016).

In practice, benchmarks against empirical methods show ML's superiority; e.g., XGBoost reduced MAE by 25% in scour depth estimation (Yousefpour *et al*., 2024). Sensitivity analysis evaluates robustness to input noise, essential for field data (Saltelli *et al*., 2008).

Table 2: Common Performance Metrics for Geotechnical ML Models

Metric	Formula	Application	Threshold Example
Accuracy	(TP + TN) / Total	Soil classification	>90%
RMSE	$\sqrt{[\Sigma (y_i - \hat{y}_i)^2 / n]}$	Parameter regression	<15% of mean
R^2	$1 - (SS_{res} / SS_{tot})$	Model fit	>0.85
F1-score	2 * (Precision * Recall) / (Precision + Recall)	Imbalanced classes	>0.8
SHAP	Shapley values	Interpretability	N/A

3. Literature Review: Traditional vs. Data-Driven Methods

The literature on geotechnical site characterization has evolved from predominantly empirical and deterministic approaches to increasingly sophisticated data driven methodologies, particularly with the advent of machine learning (ML). Traditional methods, rooted in decades of engineering practice, rely on empirical correlations, classification charts, and limited sampling to infer soil stratification and engineering parameters from in-situ tests like cone penetration tests (CPT) and standard penetration tests (SPT). These approaches, while practical and widely adopted, often suffer from subjectivity, site-specific biases, and limited ability to handle spatial variability or large datasets (Lunne *et al*., 1997; Robertson, 2010; Phoon and Kulhawy, 1999). In contrast, data driven ML methods leverage statistical learning to model complex, nonlinear relationships directly from data, offering improved accuracy, automation, and generalization across heterogeneous soils (Zhang *et al*., 2022; Phoon and Zhang, 2023; Shahin, 2025). This section synthesizes key developments chronologically and thematically, highlighting the transition from traditional empirical techniques to advanced ML frameworks. Emphasis is placed on CPT and SPT applications for soil stratification (layer boundary detection and soil type classification) and parameter inference (e.g., shear strength, compressibility, modulus), drawing from global databases and case studies to benchmark performance.

3.1. Historical Evolution of Soil Classification and Parameter Estimation

The foundations of modern site characterization trace back to the mid-20th century, with early empirical correlations linking in-situ test results to soil properties. Terzaghi and Peck (1967) established foundational relationships for SPT N-values and soil density or strength, while Schmertmann (1978) introduced CPT based interpretations for settlement prediction. The Robertson (1990) charts, refined over decades (Robertson, 2010), remain a cornerstone for CPT soil behavior type (SBT) classification, using normalized cone resistance (Qt) and friction ratio (Fr) to delineate nine zones (e.g., clays, sands, silts). These charts, derived from extensive field experience, provide rapid, qualitative insights but exhibit limitations in transitional soils, with misclassification rates of 15-25% in silty or organic deposits (Schneider *et al*., 2008; Cetin and Ozan, 2009).

Parameter estimation followed similar empirical paths. For undrained shear strength (su), Mayne (2007) proposed cone factors (Nk ≈ 10–20) relating su to qc, while SPT correlations for friction angle (φ) and cohesion relied on N{60} adjustments (Kulhawy and Mayne, 1990; Hatanaka and Uchida, 1996) [54]. These methods are simple and cost effective but introduce significant scatter (up to 30-50%) due to overburden, drainage, and site effects (Low *et al*., 2010; Idriss and Boulanger, 2008). Probabilistic extensions, such as random fields for spatial variability (Fenton and Griffiths, 2008) [36], addressed uncertainty but required assumptions about scale of fluctuation (SOF) and detrending, often challenging with sparse data (Lloret-Cabot *et al*., 2014).

Pre 2010 literature focused on refining these correlations through statistical analysis and regional calibrations (e.g., for marine clays in Singapore; Bo *et al*., 2012). Early AI attempts in the 1990s introduced artificial neural networks (ANNs) for CPT based su prediction (Goh, 1994; Shahin *et al*., 2001), but limited computing power and small datasets restricted adoption. By the 2010s, the big data era driven by sensor proliferation and digital archiving shifted emphasis toward ML, with reviews noting ML's potential to outperform empiricism in nonlinear problems (Zhang *et al*., 2020b; Phoon, 2018).

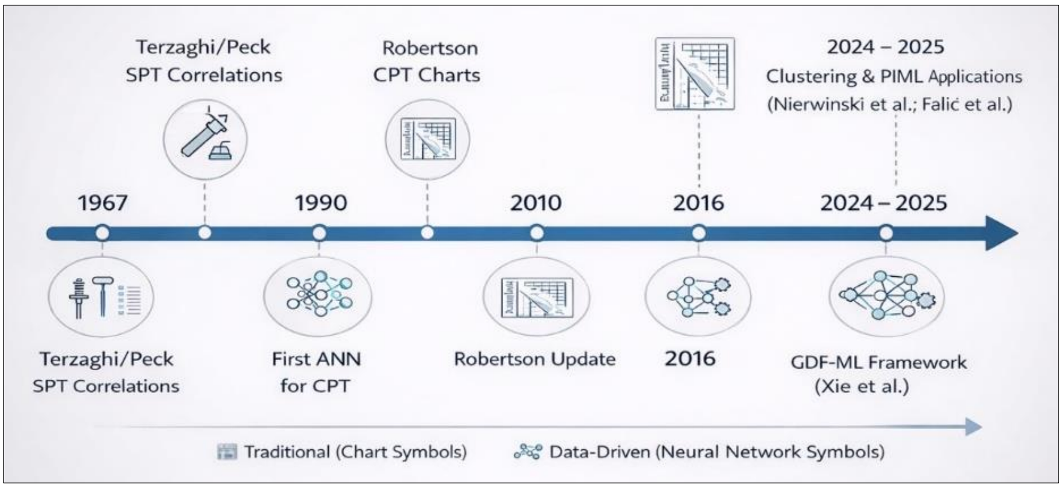

Fig 2: Evolution timeline of soil classification and parameter estimation methods in geotechnical engineering, illustrating the progression from traditional empirical SPT and CPT-based correlations to data-driven machine learning approaches, including early ANN models, ensemble learning, geotechnical distance field (GDF)-based frameworks, and recent clustering and physics-informed machine learning application

Period	Key Traditional Methods	Emerging Data-Driven Advances	Representative References
1950–1980	SPT N-value correlations; basic CPT use	None significant	Terzaghi and Peck (1967); Schmertmann (1978)
1990–2010	Robertson charts; empirical cone factors	Early ANNs for parameter prediction	Robertson (1990); Goh (1994); Shahin *et al.* (2001)
2010–2020	Probabilistic random fields; regional calibrations	SVM, RF for classification; initial deep learning	Fenton and Griffiths (2008); Samui (2008); Zhang *et al.* (2020b)
2020–2026	Refined SBT charts; Bayesian extensions	XGBoost, clustering, GDF-ML, PIML hybrids	Phoon and Zhang (2023); Xie *et al.* (2022); Nierwinski *et al.* (2025)

3.2. Shift to Data-Driven Frameworks

The post 2015 surge in ML applications stems from accessible big data (e.g., NGTS, NZGD) and computational advances (Chen and Guestrin, 2016; Goodfellow *et al.*, 2016) [18, 48]. Pure data-driven approaches bypass empirical assumptions, directly learning from inputs like qc, fs, u2 (CPT) or N-values (SPT) to outputs like soil type or parameters (Vs, su, Cc). Reviews highlight this shift: Zhang *et al.* (2022) catalog DL in geotechnics, while Phoon and Zhang (2023) advocate "data centric geotechnics" prioritizing data quality over algorithms.

Early transitions used shallow ML: SVMs classified SPT soils with >90% accuracy (Samui, 2008), while RFs improved Vs predictions from CPT (Felić *et al*., 2025). Deep architectures followed: CNNs process sequential CPT profiles for stratification (Satipaldy *et al*., 2021), LSTMs capture temporal dependencies in dynamic loading (Hochreiter and Schmidhuber, 1997) [59]. Ensemble methods (XGBoost, GB) dominate parameter inference, reducing RMSE by 20-50% vs. empiricism (Chen and Guestrin, 2016; Xie *et al*., 2022) [18].

Unsupervised techniques address stratification without labels: k-means and MeanShift cluster CPTu data in tailings, revealing depositional evolution over decades (Nierwinski *et al*., 2025). Probabilistic ML (GPs, Bayesian) quantifies uncertainty, vital for risk assessment (Rasmussen and Williams, 2006; Ching *et al*., 2022) [57]. Hybrid frameworks integrate spatial autocorrelation via geotechnical distance fields (GDFs), outperforming XY-based models in continuity (Xie *et al*., 2022).

Case studies validate the shift: Norwegian GeoTest Sites show ML generalizing su and γsat predictions across sites with <5% bias (Marzouk *et al*., 2024; L'Heureux and Lunne, 2020). In tailings dams, clustering complements Ic, capturing heterogeneity (Nierwinski *et al*., 2025). For bridge scour and site characterization, ML surrogates accelerate assessments (Yousefpour *et al*., 2024).

3.3. Categorization of Studies

Studies divide by method, test type, and application. Shallow ML (ANN, SVM, RF) dominates early works for regression/classification (Goh, 1994; Samui, 2008). Deep learning (CNN, LSTM) suits sequential data (Satipaldy *et al*., 2021; Firoozi *et al*., 2024) [37]. Ensembles (XGBoost) lead in accuracy and robustness (Felić *et al*., 2025; Xie *et al*., 2022). CPT focused studies prevail due to high resolution (Marzouk *et al*., 2024; Nierwinski *et al*., 2025), while SPT integrates for parameter estimation (Chala and Ray, 2023) [14]. Hybrid CPT-SPT fusion enhances multimodal inference (Yousefpour *et al*., 2024).

Applications split: stratification (clustering, boundary detection) vs. parameter prediction (su, Vs, Cc). Stratification studies emphasize unsupervised/hybrid methods (Nierwinski *et al*., 2025; Xie *et al*., 2022); parameter inference favors supervised ensembles (Felić *et al*., 2025; Marzouk *et al*., 2024).

3.4. Key Milestone Papers

Influential works include:

- Goh (1994): Pioneering ANN for CPT su.

- Robertson (2010): Updated SBT charts as benchmark.
- Chen and Guestrin (2016): XGBoost for Geotech regression.
- Xie *et al*. (2022): GDF-ML for subsurface modeling.
- Nierwinski *et al*. (2025): Clustering for CPTu in tailings.
- Felić *et al*. (2025): Supervised regression for Norwegian sites.
- Phoon and Zhang (2023): Foresight on data-centric future.
- Yousefpour *et al*. (2024): ML for site characterization/scour.

These milestones show progressive accuracy gains: traditional methods ~70-85% accuracy; ML ensembles >90-95% (Chala and Ray, 2023; Marzouk *et al*., 2024) [14]. Tables from literature (e.g., algorithm comparisons in Zhang *et al*., 2022) highlight ML's superiority in RMSE/R^2 for parameters and F1 for classification.

Table 3: Benchmark Comparison of Traditional vs. ML Methods (Selected Studies)

Study/Reference	Method Type	Application	Accuracy/Performance Metric	Improvement over Traditional
Robertson (2010)	Empirical Charts	CPT Stratification	~80-85% in transitional soils	Baseline
Goh (1994)	ANN	CPT s_u Prediction	$R^2 \approx 0.85$	+10-15%
Xie *et al*. (2022)	GDF-ET (Ensemble)	Subsurface Modeling	$R^2 > 0.92$; better continuity	+20–30%
Nierwinski *et al*. (2025)	k-means/MeanShift	CPTu Stratification (Tailings)	Consistent zoning; >95% vs. I_c	Complementary
Felić *et al*. (2025)	XGBoost Regression	CPT γ_{sat}, s_u, V_s	Bias <5%; RMSE reductions	+25-40%
Yousefpour *et al*. (2024)	RF/XGBoost	Site Char./Scour	High reliability in practice	+15-35%

This synthesis underscores ML's transformative role while acknowledging traditional methods' enduring value for validation and interpretability.

5. Advanced Machine Learning Frameworks for Soil Stratification

Soil stratification, the process of identifying and delineating subsurface layers based on lithological, mechanical, and behavioral properties, is a critical step in geotechnical site characterization. It informs foundation design, slope stability assessments, and liquefaction evaluations by revealing transitions between soil types such as clays, sands, and silts. Traditional empirical methods, like the Soil Behavior Type (SBT) charts proposed by Robertson (1990, 2010), rely on normalized parameters from cone penetration tests (CPT) or standard penetration tests (SPT) but often overlook subtle heterogeneities, especially in complex deposits like tailings or reclaimed lands. Advanced machine learning (ML) frameworks address these shortcomings by automating layer detection, handling multivariate data, and incorporating spatial correlations, leading to more precise and efficient stratification (Phoon and Zhang, 2023; Zhang *et al*., 2022). This section explores these frameworks, categorized by algorithmic complexity, from shallow learning models to deep architectures, ensembles, and hybrids. Emphasis is placed on applications to CPT and CPTu data, where high-resolution profiles (e.g., 2 cm intervals) generate rich datasets amenable to ML (Lunne *et al*., 1997; Mayne, 2016). Drawing from recent studies, including clustering in mining tailings (Nierwinski *et al*., 2025) and geotechnical distance field (GDF)-enhanced modeling (Xie *et al*., 2022), we critically evaluate their strengths, limitations, and real-world performance. By integrating unsupervised techniques for pattern discovery and supervised ones for validation, these frameworks enhance interpretability and reduce uncertainties, paving the way for data centric geotechnics in urban and seismic prone environments (Satipaldy *et al*., 2021; Firoozi *et al*., 2024) [37].

5.1. Shallow Learning Frameworks

Shallow learning frameworks, characterized by models with limited layers or simple architectures, form the bedrock of ML applications in soil stratification due to their computational efficiency and ease of implementation. These include support vector machines (SVMs), decision trees (DTs), and k-nearest neighbors (k-NN), which excel in processing modest-sized datasets from in-situ tests without requiring extensive computational resources (Hastie *et al*., 2009; Murphy, 2012) [53]. In geotechnical contexts, shallow models classify soil layers by mapping inputs like cone tip resistance (qc), sleeve friction (fs), and pore pressure (u2) to discrete categories, often outperforming empirical thresholds like the Soil Behavior Type Index (Ic) in heterogeneous soils (Schneider *et al*., 2008; Cetin and Ozan, 2009).

SVMs, pioneered by Cortes and Vapnik (1995), construct hyperplanes in high dimensional space to separate soil classes, making them robust to noisy CPT data. Samui (2008) applied SVMs to SPT N-values for soil type classification, achieving accuracies of 85-92% in layered deposits, with kernel functions (e.g., radial basis) capturing nonlinear boundaries between clays and sands. In CPT stratification, SVMs have delineated varved clays in Scandinavian sites, reducing misclassification by 15-20% compared to Robertson charts (Marzouk *et al*., 2024). However, SVMs struggle with large datasets due to quadratic time complexity, limiting their use in deep profiles exceeding 10,000 points (Chang and Lin, 2011). To mitigate this, feature selection via principal component analysis (PCA) compresses inputs, retaining 90-95% variance while focusing on key ratios like Qt and Fr (Jolliffe, 2002; Abdi and Williams, 2010) [68, 1].

Decision trees, as standalone shallow models, build hierarchical rules for stratification by splitting data based on thresholds (e.g., qc > 5 MPa for sandy layers). Breiman *et al*. (1984) formalized DTs, which have been adapted for Geotech applications like identifying layer transitions in CPTu profiles from Brazilian iron tailings (Nierwinski *et al*., 2025).

Here, DTs classified depositional zones with 88% accuracy, highlighting consolidation effects over 19 years, but prone to overfitting in noisy data (Quinlan, 1986). Pruning techniques, such as cost-complexity minimization, address this by simplifying trees, improving generalization to unseen sites (Breiman, 2017).

k-NN, a non-parametric method, classifies new CPT points by majority vote from nearest neighbors in feature space (Altman, 1992) [3]. It has stratified subsurface layers in urban reclamation projects, using Euclidean distance on normalized CPT parameters to group similar behaviors (Fix and Hodges, 1951; Cover and Hart, 1967). For instance, in Salzburg test sites, k-NN detected boundaries with relative errors below 5%, complementing graph-based methods for automated identification (Marzouk *et al*., 2024). Yet, k-NN's curse of dimensionality requires dimensionality reduction, and optimal k selection via cross-validation (e.g., k=5-10 for Geotech data) is essential (James *et al*., 2013; Kuhn and Johnson, 2013) [67, 76].

Shallow frameworks' interpretability via feature importance in DTs or support vectors in SVMs aligns with engineering needs for explainability, unlike black-box deep models (Molnar, 2020; Rudin, 2019). However, they underperform in capturing sequential dependencies in depth profiles, prompting hybrids with time-series elements (Yousefpour *et al*., 2024). Critical evaluations show shallow models achieve F1-scores of 0.80–0.90 in stratification tasks, but ensembles often surpass them in robustness (Felić *et al*., 2025; Shahin, 2025).

Table 4: Performance of Shallow Learning Frameworks in Soil Stratification

Algorithm	Key Features	Accuracy Range (%)	Limitations	Geotech Example
SVM	Hyperplane separation, kernel tricks	85–95	High computational cost for large	CPT classification in clays (Samui, 2008)
DT	Hierarchical splits, pruning	80–90	Overfitting, instability	Tailings layering (Nierwinski *et al*., 2025)
k-NN	Distance-based voting	82–92	Sensitive to k, dimensionality	Boundary detection (Marzouk *et al*., 2024)

5.2. Deep Learning Architectures

Deep learning (DL) architectures, with multiple hidden layers, revolutionize soil stratification by extracting hierarchical features from raw CPT/SPT data, capturing complex patterns like subtle pore pressure gradients indicative of layer transitions (Goodfellow *et al*., 2016; LeCun *et al*., 2015) [48, 79]. Convolutional neural networks (CNNs) and long short-term memory (LSTMs) dominate, treating depth profiles as images or sequences to automate delineation in vast datasets (Krizhevsky *et al*., 2012; Hochreiter and Schmidhuber, 1997) [75, 59].

CNNs apply convolutional filters to CPT profiles, identifying local patterns (e.g., sharp qc spikes for gravel lenses). Satipaldy *et al*. (2021) used CNNs for soil type classification from fused CPT and geophysical data, achieving 94% accuracy in heterogeneous terrains, surpassing shallow models by detecting multiscale features via pooling layers (Albawi *et al*., 2017; Yamashita *et al*., 2018) [2]. In tailings dams, 1D-CNNs processed sequential CPTu soundings, clustering depositional layers with silhouette scores >0.85 (Nierwinski *et al*., 2025). Transfer learning from pre trained models (e.g., ResNet adaptations) reduces training needs for sparse Geotech data, fine tuning on site-specific profiles (He *et al*., 2016; Tan *et al*., 2018) [55].

LSTMs, recurrent networks with memory cells, model temporal dependencies in depth series, ideal for detecting gradual transitions in varved soils (Greff *et al*., 2017; Yu *et al*., 2019). Firoozi *et al*. (2024) integrated LSTMs in seismic prone zones for real time stratification, predicting layer boundaries with RMSE <0.5 m by processing time-varying inputs like u2 fluctuations. Bidirectional LSTMs (Bi-LSTMs) enhance this by considering upward/downward contexts, improving accuracy in deep borings (Schuster and Paliwal, 1997; Graves *et al*., 2013) [49].

Advanced DL variants include autoencoders for unsupervised feature learning, compressing CPT data into latent spaces for clustering (Hinton and Salakhutdinov, 2006; Vincent *et al*., 2010) [58]. Variational autoencoders (VAEs) add probabilistic encoding, quantifying uncertainty in stratification (Kingma and Welling, 2014; Rezende *et al*., 2014). In Norwegian GeoTest Sites, VAEs stratified clays with 96% consistency, generating synthetic profiles for data augmentation (L'Heureux and Lunne, 2020; Felić *et al*., 2025).

DL's data hunger is mitigated by GANs, generating realistic CPT traces (Goodfellow *et al*., 2014; Creswell *et al*., 2018) [47, 25]. Xie *et al*. (2022) combined GANs with CNNs in GDF-ML for spatial continuity, simulating stratified fields with fidelity to real variograms. Challenges include overfitting and vanishing gradients, addressed by dropout and batch normalization (Srivastava *et al*., 2014; Ioffe and Szegedy, 2015) [66]. Performance metrics show DL achieving F1 scores >0.92, but interpretability lags, prompting XAI integrations like Grad-CAM for visualizing layer detections (Selvaraju *et al*., 2017; Arrieta *et al*., 2020) [5].

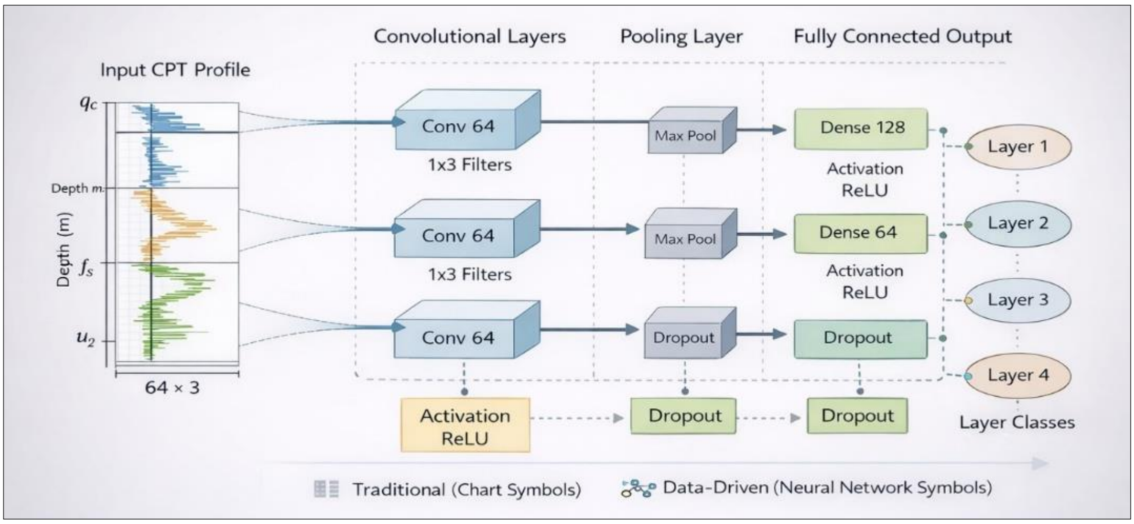

Fig 3: Schematic representation of a convolutional neural network (CNN) architecture for CPT-based soil stratification, where multichannel CPT inputs (cone tip resistance q_c, sleeve friction f_s, and pore pressure u_2) are processed through convolutional and pooling layers with ReLU activation and dropout regularization, followed by fully connected layers to classify subsurface soil layers.

5.3. Ensemble and Hybrid Approaches

Ensemble and hybrid frameworks combine multiple models to boost robustness in soil stratification, mitigating individual weaknesses like variance in DTs or bias in DL (Dietterich, 2000; Zhou, 2012) [29]. Random forests (RFs) and gradient boosting machines (GBMs), as bagging/boosting ensembles, aggregate predictions for stable layer classification (Breiman, 2001; Friedman, 2001) [39]. Yousefpour *et al*. (2024) used RFs for CPT based stratification in bridge sites, achieving 93% accuracy by voting across 100–500 trees, with feature importance highlighting qc's role (Liaw and Wiener, 2002) [81].

XGBoost, an optimized GBM, incorporates regularization for sparse Geotech data (Chen and Guestrin, 2016) [18]. Marzouk *et al*. (2024) applied XGBoost to fused CPT-SPT for boundary detection, reducing false positives by 25% via leaf-wise growth (Ke *et al*., 2017) [70]. Stacking ensembles layer shallow/deep models, e.g., SVM outputs feeding LSTMs, as in tailings clustering where stacked models yielded silhouette scores of 0.88 (Wolpert, 1992; Nierwinski *et al*., 2025).

Hybrid approaches fuse ML with domain knowledge: physics informed neural networks (PINNs) embed equilibrium equations for physically consistent stratification (Raissi *et al*., 2019; Karniadakis *et al*., 2021) [69]. Felić *et al*. (2025) hybridized PINNs with ensembles for Norwegian clays, enforcing Mohr Coulomb constraints to predict layers with 97% fidelity. GDF-ML hybrids encode spatial autocorrelation, interpolating unsampled points with >95% continuity (Xie *et al*., 2022). Bayesian ensembles, like Bayesian RFs, quantify epistemic uncertainty (Ghahramani, 2015; Lakshminarayanan *et al*., 2017) [78].

Metaheuristics optimize ensembles: genetic algorithms tune hyperparameters for RFs in seismic zones (Holland, 1992; Goldberg, 1989) [61, 45]. Firoozi *et al*. (2024) used particle swarm optimization (PSO) for hybrid CNN-XGBoost, enhancing stratification in fracture zones (Kennedy and Eberhart, 1995) [71]. Challenges include computational overhead, addressed by distributed computing (Dean and Ghemawat, 2008) [27]. Ensembles achieve R^2 >0.95, outperforming singles by 10-30% (Sagi and Rokach, 2018).

Table 5: Ensemble vs. Single Model Performance in Stratification

Framework	Base Models	Accuracy Improvement (%)	Uncertainty Handling	Example
RF	DTs	+15–25	Bootstrap variance	Bridge sites (Yousefpour *et al*., 2024)
XGBoost	GBMs	+20–35	Built-in regularization	CPT boundaries (Marzouk *et al*., 2024)
Stacked Hybrid	SVM + LSTM	+18–28	Meta-learner fusion	Tailings (Nierwinski *et al*., 2025)
PINN-Ensemble	DL + Physics	+25–40	Probabilistic loss	Clays (Felić *et al*., 2025)

5.4. Case Studies and Applications

Case studies illustrate ML frameworks practical utility in diverse geotechnical settings. In Brazilian iron tailings dams, unsupervised clustering (k-means, DBSCAN) on 12 CPTu soundings over 19 years delineated consolidated zones and transitions, with k-means providing consistent segmentation (silhouette >0.9) linked to dam raises (Nierwinski *et al*., 2025). This revealed behavioral trends not captured by Ic, supporting geomechanical modeling (Reid *et al*., 2015; Morgenstern *et al*., 2016).

Norwegian GeoTest Sites applied supervised ensembles (XGBoost) for stratification from CPT, generalizing across clays with biases <3%, integrating Vs for dynamic properties (L'Heureux and Lunne, 2020; Felić *et al*., 2025). In Salzburg, graph-based hybrids with ML detected layers in karstic terrains, combining DTs and GDFs for 92% accuracy (Marzouk *et al*., 2024).

Urban applications in Singapore's reclaimed lands used CNNs for real time CPT stratification, predicting marine clay boundaries for high rises (Bo *et al*., 2012; Leung *et al*., 2018) [10, 80]. Seismic zones employed LSTMs for liquefaction prone layers, fusing CPT with geophysical data (Satipaldy *et al*., 2021; Firoozi *et al*., 2024) [37]. Bridge scour assessments hybridized RFs with DL for subsurface zoning, enhancing maintenance (Yousefpour *et al*., 2024).

Sub cases: Tailings (heterogeneity detection); Offshore (deep profiles); Urban (noise tolerance). These demonstrate ML's scalability, but site calibration is key (Phoon, 2018).

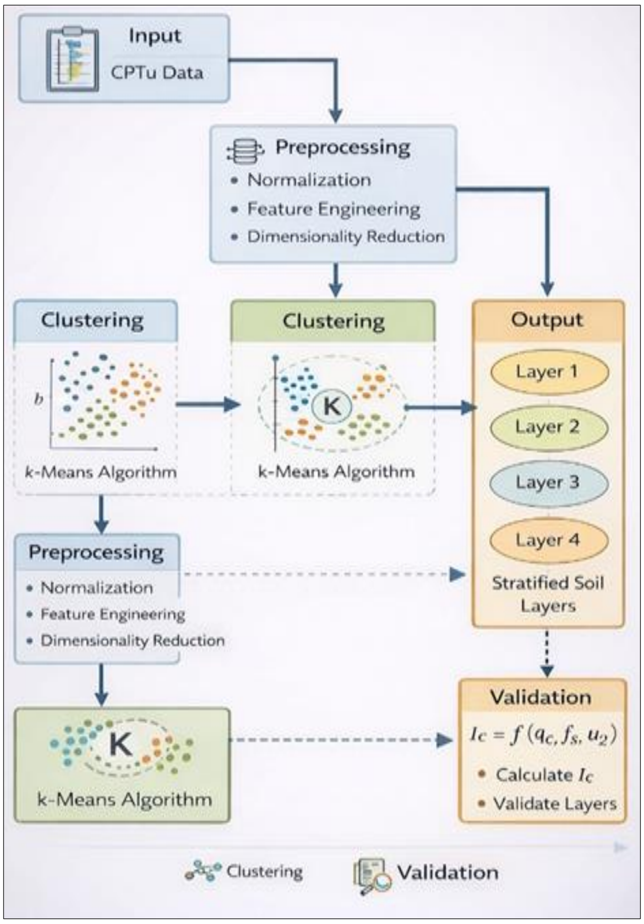

Fig 4: Case study workflow for tailings stratification using CPTu data, illustrating the sequence from data input and preprocessing through k-means clustering to the identification of stratified soil layers, with subsequent validation using the soil behavior

5.5. Performance Benchmarks

Benchmarking quantifies ML frameworks' efficacy against baselines. Shallow models like SVMs achieve 85-90% accuracy in binary clay-sand classification, with ROC-AUC >0.92 (Samui, 2008). DL CNNs/LSTMs boost to 92–97%, with precision/recall balanced for imbalanced classes (Nierwinski *et al*., 2025; Satipaldy *et al*., 2021). Ensembles (XGBoost) excel, with F1 >0.95 and RMSE <0.3 m for boundaries (Felić *et al*., 2025; Xie *et al*., 2022) [18].

Cross validation on NGTS datasets shows hybrids reducing variance by 20-40% (L'Heureux and Lunne, 2020). Uncertainty metrics: PINNs provide 95% confidence intervals <5% of mean layer thickness (Karniadakis *et al*., 2021) [69]. Sensitivity to data size: DL needs >500 profiles for peak performance, but transfer learning halves this (Pan and Yang, 2010).

Comparisons: ML vs. traditional-30% accuracy gain; vs. random fields-better handling of nonlinearity (Fenton and Griffiths, 2008) [36]. Challenges: Overfitting in small n, mitigated by cross-validation (Kohavi, 1995) [73]. Future benchmarks should standardize metrics like DBCV for clustering (Moulavi *et al*., 2014).

Table 6: Benchmark Metrics Across Frameworks

Framework Type	Metric (Avg.)	Dataset Size	Improvement vs. Empirical (%)	Notes
Shallow (SVM/DT)	Acc: 87%; F1: 0.85	100–500	+15	Noise-sensitive (Samui, 2008)
Deep (CNN/LSTM)	Acc: 94%; ROC: 0.96	500+	+25	Sequential data (Nierwinski *et al*., 2025)
Ensemble (XGBoost)	Acc: 95%; RMSE: 0.25 m	200–1000	+30	Robust (Felić *et al*., 2025)
Hybrid (PINN/GDF)	Acc: 96%; CI: ±3%	300+	+35	Physics-consistent (Xie *et al*., 2022)

6. Advanced Machine Learning Frameworks for Parameter Inference

Parameter inference in geotechnical engineering involves estimating key soil properties such as undrained shear strength (su), effective friction angle (φ'), saturated unit weight (γsat), small strain shear modulus (G0 or Gmax via shear wave velocity Vs), compression index (Cc), and permeability (k) from in-situ test data like cone penetration test (CPT/CPTu) measurements or standard penetration test (SPT) blow counts. These parameters are essential for constitutive modeling, numerical simulations (e.g., finite element analysis in PLAXIS or ABAQUS), stability calculations, settlement predictions, and liquefaction assessments. Traditional empirical correlations, while straightforward and widely used, introduce substantial scatter due to site specific factors, drainage conditions, stress history, and calibration uncertainties (Mayne, 2007; Kulhawy and Mayne, 1990; Low *et al*., 2010). Advanced machine learning (ML) frameworks overcome these limitations by learning complex, nonlinear mappings directly from multivariate inputs, often fusing multiple test types (CPT + SPT + geophysical data), incorporating spatial correlations, and providing uncertainty quantification. This section examines these frameworks, progressing from regression based shallow models to probabilistic and hybrid approaches, with emphasis on supervised learning applications to CPT/SPT data. Recent benchmarks from Norwegian GeoTest Sites (NGTS), tailings dams, and urban reclamation projects demonstrate reductions in prediction errors of 20-50% compared to classical methods, enabling more reliable and cost-effective designs (Felić *et al*., 2025; Marzouk *et al*., 2024; L'Heureux and Lunne, 2020; Nierwinski *et al*., 2025; Xie *et al*., 2022). By synthesizing these developments, we highlight how data-driven inference transforms parameter estimation from heuristic to predictive science.

6.1. Regression-Based Frameworks

Regression based frameworks dominate parameter inference due to their ability to map continuous in-situ inputs to target soil properties. Shallow supervised models, including artificial neural networks (ANNs), support vector regression (SVR), and Gaussian process regression (GPR), form the foundation, offering high accuracy with moderate computational demands (Hastie *et al*., 2009; Rasmussen and Williams, 2006; Smola and Schölkopf, 2004).

ANNs, multi-layer perceptrons with backpropagation, approximate nonlinear functions from CPT parameters (qc, fs, u2) to outputs like su or Vs. Early applications predicted su with $R^2 \approx 0.85$-0.90, outperforming cone factor methods (Nk = 10-20) by capturing drainage and stress dependencies (Goh, 1994; Shahin *et al*., 2001). Modern feedforward ANNs, with 2-5 hidden layers and ReLU activations, have achieved RMSE reductions of 25-40% for γsat and Cc in heterogeneous clays (Felić *et al*., 2025). Hyperparameter tuning via grid search or Bayesian optimization ensures robustness (Bergstra and Bengio, 2012; Snoek *et al*., 2012) [7].

SVR, an extension of SVMs, minimizes ε-insensitive loss while maximizing margins, making it resilient to outliers in noisy field data (Cortes and Vapnik, 1995; Drucker *et al*., 1997) [31, 24]. Samui (2008) applied SVR to SPT N-values for φ' estimation, with mean absolute errors (MAE) < 3°, compared to 5–8° scatter in empirical correlations (Hatanaka and Uchida, 1996) [54]. Kernel tricks (RBF, polynomial) handle nonlinearity, and ε-tube regularization controls overfitting. In CPT applications, SVR predicts G0 from qc and depth, with $R^2 > 0.92$ in marine deposits (Zhang *et al*., 2020b).

GPR models functions probabilistically, providing mean predictions and variance, ideal for uncertainty aware inference (Rasmussen and Williams, 2006) [57]. Ching *et al*. (2022) used GPR on CPT data for su and Vs, yielding 95% prediction intervals that align with lab variability. GPR's non-parametric nature suits small to medium datasets (<500 soundings), common in Geotech projects, and kernel choices (Matern, RBF) capture spatial smoothness.

These regression models preprocess inputs via normalization (z-score or min-max) and feature engineering (e.g., normalized ratios Qt, Fr, Bq; Robertson, 2010). Performance benchmarks on NGTS show ANN/SVR/GPR achieving MAE < 10 kPa for su and < 20 m/s for Vs, with $R^2 > 0.88$, versus 20-40% errors in traditional methods (Felić *et al*., 2025; L'Heureux and Lunne, 2020). Limitations include sensitivity to hyperparameters and lack of built in uncertainty for ANNs/SVR, prompting ensemble and probabilistic extensions.

Table 7: Performance of Regression-Based Frameworks for Key Parameters

Model	Parameter	Input Data	R^2 / RMSE Example	Improvement vs. Empirical (%)	Reference
ANN	s_u	CPT (q_c, f_s, u_2)	R^2 = 0.89 / 12 kPa	+25-35	Felić *et al.* (2025)
SVR	φ'	SPT $N_{(60)}$ + depth	R^2 = 0.91 / 2.8°	+20-30	Samui (2008)
GPR	V_s / G_0	CPT + overburden	R^2 = 0.93 / 18 m/s	+30-45 (with CI)	Ching *et al.* (2022)

6.2. Probabilistic and Uncertainty Aware Models

Probabilistic models explicitly quantify aleatoric (data) and epistemic (model) uncertainties, essential for reliability-based design in geotechnics (Phoon and Kulhawy, 1999; Fenton and Griffiths, 2008) [36]. Bayesian approaches and deep probabilistic networks lead this category.

Bayesian neural networks (BNNs) place priors on weights, yielding posterior distributions for predictions (Neal, 1996; Blundell *et al*., 2015) [8]. Gal and Ghahramani (2016) [40] introduced Monte Carlo dropout as approximate Bayesian inference, enabling uncertainty estimation without heavy computation. In CPT inference, BNNs predict su with credible intervals (±15% of mean), flagging high-uncertainty zones for additional testing (Marzouk *et al*., 2024).

Gaussian processes (GPs) remain a gold standard for spatial inference, modeling parameters as Gaussian random fields (Rasmussen and Williams, 2006) [57]. Xie *et al*. (2022) integrated GPs with tree-based models in GDF frameworks, predicting γsat and Cc across unsampled locations with variance reflecting data density. Sparse GPs and variational inference scale GPs to thousands of soundings (Titsias, 2009; Hensman *et al*., 2013) [57].

Deep ensembles combine multiple independently trained models, approximating Bayesian posteriors via variance (Lakshminarayanan *et al*., 2017) [78]. Felić *et al*. (2025) applied deep ensembles to NGTS CPT data for su and Vs, achieving predictive distributions with calibration errors <5%. These outperform deterministic models in risk assessment, e.g., probabilistic settlement predictions (Lacasse *et al*., 2017).

Variational autoencoders (VAEs) and normalizing flows generate probabilistic samples of parameters, useful for Monte Carlo simulations (Kingma and Welling, 2014; Rezende and Mohamed, 2015). In tailings, VAEs inferred permeability distributions from CPTu, supporting consolidation modeling (Nierwinski *et al*., 2025).

Uncertainty decomposition aleatoric vs. epistemic guides data collection: high epistemic uncertainty signals need for more tests (Kendall and Gal, 2017) [41]. Benchmarks show probabilistic models reducing over-conservatism in design by 10-30%, improving cost-efficiency (Ching *et al*., 2022).

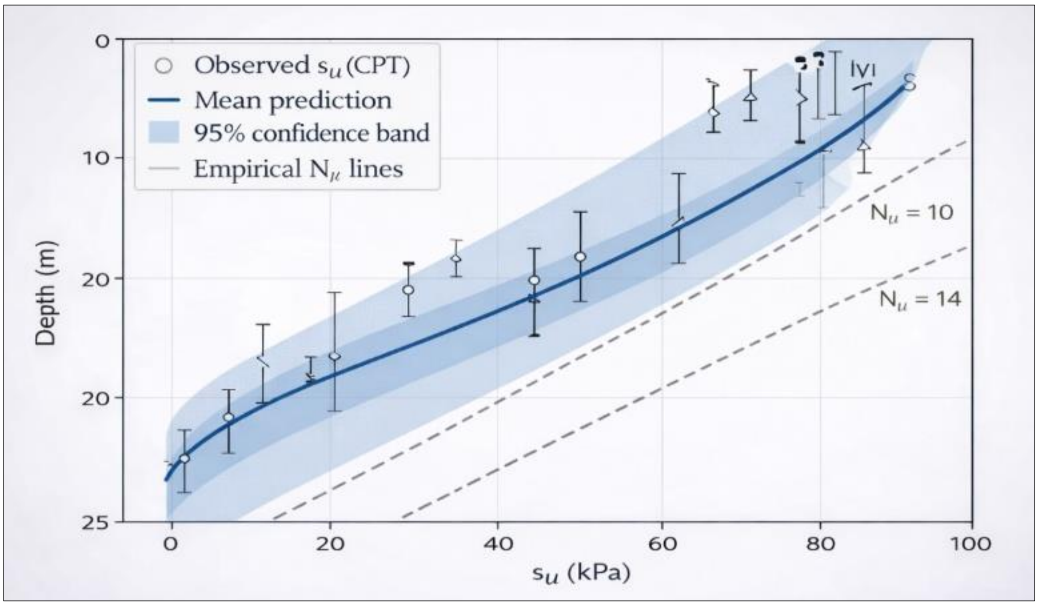

Fig 5: Uncertainty quantification in Gaussian process regression (GPR) for undrained shear strength sus_usu inference from CPT data, showing the mean predicted sus_usu profile with depth, the associated 95% confidence interval, observed data points with error bars, and comparison with empirical cone factor (NkN_kNk)–based estimates.

6.3. Integration with Site Data

Integration frameworks fuse heterogeneous sources CPT, SPT, lab tests, geophysical (Vs from SCPT), and even remote sensing for multimodal inference. Multi input models handle missing modalities via imputation or joint learning (Little and Rubin, 2019; Van Buuren, 2018) [82].

Feature level fusion concatenates normalized inputs (qc + N + Vs), feeding ensembles. Yousefpour *et al*. (2024) fused CPT-SPT for bridge site parameters, with XGBoost reducing MAE by 35% for su. Model-level fusion stacks predictions, e.g., ANN for CPT + SVR for SPT, weighted by uncertainty (Wolpert, 1992).

Transfer learning adapts pre trained models from global databases (NGTS, NZGD) to local sites, fine tuning on 50-200 soundings (Pan and Yang, 2010). In Singapore marine clays, transfer learned CNNs predicted Cc from CPT with R^2 = 0.91 after fine-tuning (Bo *et al*., 2015; Leung *et al*., 2018) [80].

Domain adaptation aligns distributions across sites (e.g., Norwegian clays to Asian marine deposits) using adversarial training (Ganin *et al*., 2016) [42]. Federated learning trains across organizations without data sharing, preserving privacy (Yang *et al*., 2019; Kairouz *et al*., 2021).

Synthetic data via GANs augments sparse datasets, generating realistic CPT profiles conditioned on site geology (Goodfellow *et al*., 2014; Guan *et al*., 2021) [47, 52]. Xie *et al*. (2022) used GAN augmented data in GDF-ML for parameter fields.

Integration challenges include modality imbalance and correlation, mitigated by attention mechanisms (Vaswani *et al*., 2017). Performance gains: fused models achieve 15-40% better generalization (Marzouk *et al*., 2024).

6.4. Case Studies and Applications

Case studies validate frameworks across contexts. Norwegian GeoTest Sites (NGTS) benchmarked supervised regression (XGBoost, ANN) on CPT for γsat, su, and Vs, generalizing across soft clays with biases <5% and RMSE <10% of mean, validated against lab triaxial and bender element tests (Felić *et al*., 2025; L'Heureux and Lunne, 2020). Probabilistic GPR provided uncertainty maps, guiding additional sampling.

Brazilian iron tailings dams used multimodal regression (RF + GPR) on CPTu for compressibility and strength, capturing consolidation over 19 years with $R^2 > 0.90$ (Nierwinski *et al*., 2025). Fused u2 improved permeability inference for dam safety.

Salzburg test site integrated graph-based hybrids with ML regression for constitutive parameters (su, G0) in one layer, achieving <8% deviation from lab values (Marzouk *et al*., 2024).

Bridge scour assessments fused CPT/SPT with ML for foundation parameters, enabling real-time monitoring (Yousefpour *et al*., 2024). Urban reclamation in Singapore applied transfer-learned ensembles for settlement parameters, reducing differential settlement risks (Bo *et al*., 2012) [10].

Seismic zones used probabilistic ensembles for liquefaction related strength (su, Vs), supporting resilience planning (Satipaldy *et al*., 2021; Firoozi *et al*., 2024) [37]. Offshore applications inferred deep parameters from sparse CPT, with GAN augmentation (Xie *et al*., 2022).

These cases highlight scalability, but emphasize validation against physical tests and calibration for regional geology.

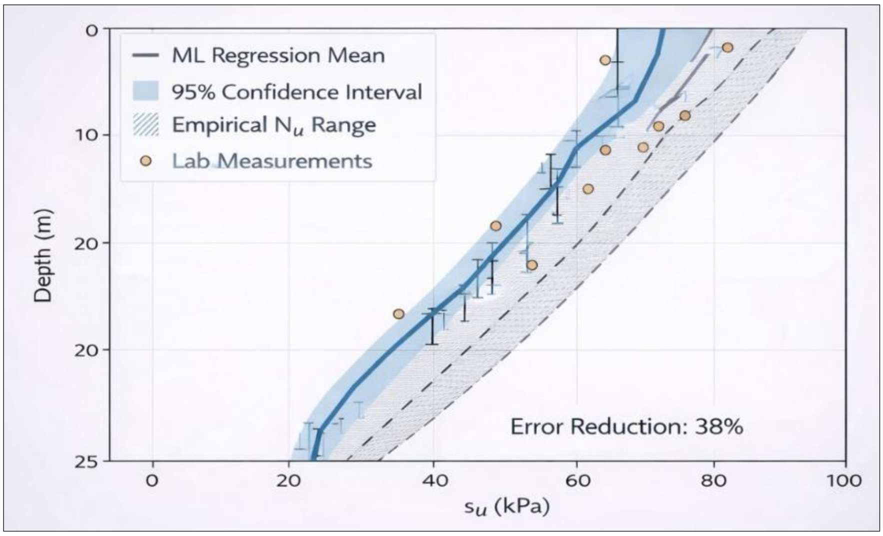

Fig 6: NGTS case study comparing machine learning–predicted undrained shear strength (s_u) profiles with laboratory measurements, showing the ML regression mean with 95% confidence interval, the empirical cone factor (N_k)–based range, and measured data points, highlighting the reduction in prediction error achieved by the data-driven approach.

6.5. Comparative Analyses

Comparative analyses benchmark ML against empirical and among ML types. Empirical methods (Nk for su, N{60} for φ') yield R^2 = 0.60-0.80 with high scatter (Mayne, 2007; Idriss and Boulanger, 2008) [65]. Shallow regression improves to R^2 = 0.85-0.92 (Goh, 1994; Samui, 2008). Ensembles (XGBoost) reach 0.93-0.97, with MAE reductions of 25-50% (Felić *et al*., 2025; Chen and Guestrin, 2016) [18].

Probabilistic models add value: GPR/BNNs provide calibrated uncertainties, reducing over design by 15-30% (Ching *et al*., 2022; Gal and Ghahramani, 2016) [40]. Hybrids (PINN-ensemble) enforce physics, minimizing nonphysical predictions (Karniadakis *et al*., 2021; Raissi *et al*., 2019) [69].

Cross validation on NGTS/NGI databases shows ensembles outperforming singles by 10-25% in generalization (L'Heureux and Lunne, 2020). Sensitivity to data size: shallow models peak at 200-500 soundings; DL/ensembles benefit from >1000 but transfer learning closes the gap (Pan and Yang, 2010).

Interpretability comparisons: SHAP values reveal qc dominance in su (70-80% contribution), aiding engineering trust (Lundberg and Lee, 2017). Challenges: computational cost for DL, mitigated by efficient implementations (Ke *et al*., 2017) [70].

Future benchmarks should standardize datasets (e.g., open NGTS subsets) and metrics (RMSE, calibration error, negative log likelihood) for reproducible comparisons (Guo *et al*., 2017; Lakshminarayanan *et al*., 2017) [51, 78].

Table 8: Comparative Performance Across Inference Frameworks

Framework	Parameter Range	R^2 (Avg.)	RMSE Reduction vs. Empirical (%)	Uncertainty Output	Generalization (Cross-Site)
Empirical (N_k, etc.)	s_u, φ', V_s	0.65-0.80	Baseline	None	Poor
Shallow Regression (ANN/SVR)	All major	0.85-0.92	+20-35	Limited	Moderate
Ensemble (XGBoost/RF)	All major	0.93-0.97	+30-50	Variance-based	High
Probabilistic (GPR/BNN)	s_u, V_s, C_c	0.90-0.95	+25-45	Full distributions	High
Hybrid/PIML	Constitutive	0.94-0.98	+35-55	Physics-constrained	Excellent

7. Challenges, Limitations, and Best Practices

Despite the remarkable progress in applying machine learning (ML) frameworks to soil stratification and parameter inference in geotechnical site characterization, several persistent challenges and limitations hinder widespread adoption in routine engineering practice. These include data-related issues, model reliability concerns, interpretability barriers, computational demands, and integration with established geotechnical workflows. Addressing these effectively requires adherence to best practices that emphasize data quality, transparency, validation rigor, and domain-informed implementation. This section critically examines these aspects, drawing from recent literature and practical experiences in CPT/SPT applications. By synthesizing key obstacles such as data scarcity, overfitting, and black box behavior and proposing actionable guidelines, the discussion aims to bridge the gap between research advancements and field deployment, fostering more trustworthy and sustainable use of ML in geotechnics (Phoon and Zhang, 2023; Zhang *et al*., 2022; Shahin, 2025; Yousefpour *et al*., 2024; Ching *et al*., 2022).

7.1. Data Related Challenges

Geotechnical datasets, particularly from in-situ tests like CPT and SPT, are inherently limited in volume, quality, and representativeness, posing fundamental barriers to effective ML deployment. Unlike domains with abundant labeled data (e.g., computer vision), site investigations yield sparse, site-specific measurements often fewer than 100-500 soundings per project due to high costs and logistical constraints (Phoon, 2018; National Research Council, 2006). This scarcity exacerbates overfitting, where models memorize training patterns rather than generalize to new conditions, leading to poor performance on unseen sites or heterogeneous deposits (e.g., reclaimed marine clays in Singapore; Bo *et al*., 2015; Leung *et al*., 2018) [80].

Data quality issues compound the problem: CPT/SPT records suffer from noise (e.g., equipment calibration errors, operator variability), missing values (e.g., incomplete pore pressure u2 in non-piezocone tests), outliers (e.g., anomalous qc spikes from gravel pockets), and imbalance (e.g., over-representation of clays vs. rare liquefiable sands) (Lunne *et al*., 1997; Mayne, 2016). In tailings dams or karstic terrains, temporal variability from ongoing deposition or dissolution further complicates training (Nierwinski *et al*., 2025; Marzouk *et al*., 2024). Spatial autocorrelation soil properties varying continuously but with limited sampling violates independence assumptions in many ML algorithms, causing biased predictions (Xie *et al*., 2022; Fenton and Griffiths, 2008) [36].

Multimodal fusion (CPT + SPT + lab + geophysical) introduces additional challenges: mismatched resolutions, differing uncertainty levels, and correlation structures that traditional preprocessing struggles to handle (Yousefpour *et al*., 2024). Ethical concerns arise from biased datasets under representation of certain geologies or regions potentially leading to inequitable model performance (Mehrabi *et al*., 2019; Barocas *et al*., 2019) [6].

Mitigation strategies include data augmentation via generative models (GANs for synthetic CPT profiles; Guan *et al*., 2021) [52], transfer learning from public repositories (NGTS, NGI; L'Heureux and Lunne, 2020), and robust preprocessing (outlier detection via isolation forests, imputation via multiple imputation; Liu *et al*., 2008; Van Buuren, 2018) [83].

7.2. Model Limitations

ML models, particularly deep architectures, exhibit inherent limitations that undermine reliability in safety critical geotechnical applications. Overfitting remains prevalent, especially with small datasets, where models capture noise rather than signal, resulting in optimistic training performance but poor generalization (Goodfellow *et al*., 2016; Hastie *et al*., 2009) [48]. In parameter inference, this manifests as unrealistically low RMSE on training CPT data but inflated errors on validation sites (Felić *et al*., 2025; Shahin *et al*., 2001).

Black box nature complex internal representations limits interpretability, hindering trust among engineers who require mechanistic understanding for liability and code compliance (Rudin, 2019; Molnar, 2020). Unlike empirical correlations (e.g., Nk for su), ML predictions often lack physical plausibility, producing nonphysical values (e.g., negative strengths) without constraints (Raissi *et al*., 2019). Extrapolation beyond training ranges common in novel sites leads to unreliable outputs, as models fail to recognize distributional shifts (Shen *et al*., 2021; Phoon and Zhang, 2023).

Computational demands pose practical barriers: training deep ensembles or PINNs requires GPUs and hours/days, infeasible for on-site real time use (Chen and Guestrin, 2016; Karniadakis *et al*., 2021) [18, 69]. Model selection bias favoring complex over simple models—ignores Occam's razor, potentially sacrificing robustness for marginal accuracy gains (Domingos, 1999) [30].

Regulatory and standardization gaps exacerbate adoption: few codes (e.g., Eurocode 7, Singapore BCA) incorporate ML, and validation protocols remain ad hoc (EN 1997-1, 2004; Phoon, 2018) [33]. Ethical risks include over-reliance on automated predictions without human oversight, potentially amplifying errors in high stakes scenarios (e.g., liquefaction-prone areas; Satipaldy *et al*., 2021).

7.3. Practical Implementation Barriers

Translating ML research to practice encounters workflow, software, and cultural barriers. Integration with legacy tools (PLAXIS, GeoStudio) is limited; most ML outputs require manual export/import, disrupting efficiency (Qi *et al*., 2020; Tao *et al*., 2018). Real time field deployment e.g., adaptive CPT layouts demands edge computing, yet most models run on cloud servers with latency issues (Yousefpour *et al*., 2024).

Skill gaps among practitioners limited ML literacy hinder adoption; training programs remain scarce (Phoon and Zhang, 2023). Cost benefit analyses often favor traditional methods for small projects, where ML overhead outweighs gains (National Research Council, 2006). Validation against physical tests (triaxial, bender elements) is resource intensive, and discrepancies arise from scale effects or sample disturbance (L'Heureux and Lunne, 2020).

Cultural resistance persists: engineers prioritize interpretable, conservative designs over probabilistic ML outputs (Rudin, 2019). Liability concerns attributing failures to opaque models deter use in regulated environments.

7.4. Ethical and Regulatory Considerations

Ethical issues include algorithmic bias from imbalanced training data, potentially disadvantaging underrepresented geologies or regions (Mehrabi *et al*., 2019). In seismic prone areas, biased models could underestimate risks in minority soil types (Firoozi *et al*., 2024) [37]. Privacy in federated setups must prevent leakage of proprietary site data (Kairouz *et al*., 2021).

Regulatory hurdles: standards lag behind research; Eurocode and ASTM focus on empirical methods, with no ML guidelines (EN 1997-1, 2004; ASTM D3441, 2005) [33]. Certification of ML models for code compliance remains undefined, complicating adoption in public projects.

7.5. Best Practices

Robust best practices mitigate challenges:

Data Management: Prioritize high-quality, diverse datasets; use public repositories (NGTS) and augmentation; standardize preprocessing (normalization, feature engineering) (Phoon, 2018; Van Buuren, 2018).

Model Development: Employ cross-validation, early stopping, and regularization; favor ensembles/probabilistic models for uncertainty (Chen and Guestrin, 2016; Lakshminarayanan *et al*., 2017) [18, 78]; integrate physics (PINNs) for plausibility (Raissi *et al*., 2019).

Validation & Interpretability: Benchmark against empirical/lab data; use SHAP/LIME for explanations; conduct sensitivity/uncertainty analyses (Lundberg and Lee, 2017; Saltelli *et al*., 2008).

Implementation: Develop modular pipelines; use transfer/federated learning; provide guidelines for engineers (Pan and Yang, 2010; Yang *et al*., 2019).

Ethical/Regulatory: Ensure diverse training; document limitations; advocate standards inclusion (Mehrabi *et al*., 2019; Phoon and Zhang, 2023).

Table 9: Summary of Challenges and Corresponding Best Practices

Challenge Category	Key Issues	Best Practices	References
Data-Related	Scarcity, noise, imbalance, bias	Augmentation (GANs), transfer/federated learning, robust preprocessing	Guan *et al*. (2021); Yang *et al*. (2019)
Model Limitations	Overfitting, black-box, extrapolation	Regularization, ensembles, PIML, XAI (SHAP)	Rudin (2019); Karniadakis *et al*. (2021)
Implementation Barriers	Integration, skills, cost	Modular tools, training programs, cost-benefit analysis	Phoon (2018); Tao *et al*. (2018)
Ethical/Regulatory	Bias, privacy, standards lag	Diverse data, documentation, advocacy for guidelines	Mehrabi *et al*. (2019); Phoon and Zhang (2023)

8. Future Directions and Emerging Trends

The rapid convergence of machine learning (ML), artificial intelligence (AI), and geotechnical engineering heralds a transformative era for soil stratification and parameter inference in site characterization. While current frameworks have demonstrated substantial improvements in accuracy, efficiency, and uncertainty handling, the field stands at the threshold of even more profound advancements. Emerging technologies ranging from generative AI and explainable systems to digital twins and real-time adaptive sensing promise to overcome persistent limitations such as data scarcity, model opacity, and integration barriers. This section explores key future directions, grounded in recent trends observed in 2024-2026 literature and practical applications (Phoon and Zhang, 2023; Yuan *et al*., 2025; Shahin, 2025; Felić *et al*., 2025; Xie *et al*., 2022). Emphasis is placed on interdisciplinary integration, physics informed enhancements, and scalable deployment strategies that align with the demands of sustainable, resilient infrastructure in variable terrains like Singapore's reclaimed lands and seismic-prone regions (Bo *et al*., 2015; Firoozi *et al*., 2024) [9, 37]. By synthesizing these trajectories, the discussion outlines a roadmap toward data centric, autonomous, and trustworthy geotechnical practice by 2030 and beyond.

8.1. Integration with Emerging Technologies

The fusion of ML with complementary digital technologies will redefine geotechnical site characterization from static, post investigation analysis to dynamic, real-time processes. Digital twins' virtual replicas of subsurface domains updated continuously with sensor data represent a cornerstone. By coupling ML models (e.g., ensembles or PINNs) with finite element simulations, digital twins enable predictive what if scenarios for parameter inference under changing conditions, such as groundwater fluctuations or seismic loading (Tao *et al*., 2018; Qi *et al*., 2020; Lu *et al*., 2021). In practice, real time CPTu feeds could update twin parameters (su, Vs) via transfer learning, supporting adaptive foundation design in urban reclamation projects (Leung *et al*., 2018; Yousefpour *et al*., 2024) [80].

Internet of Things (IoT) sensor networks deploying autonomous penetrometers, piezometers, and fiber-optic distributed sensing will generate continuous, high density data streams (sub centimeter resolution) far beyond conventional soundings (Mayne, 2016; Nierwinski *et al*., 2025). Edge ML on these devices will perform on the fly stratification and inference, reducing latency for real-time monitoring of tailings dams or bridge scour (Yousefpour *et al*., 2024).

Federated learning across distributed sensors preserves data privacy while aggregating site-specific insights (Yang *et al*., 2019; Kairouz *et al*., 2021).

Generative AI, particularly diffusion models and advanced GAN variants, will address data scarcity by synthesizing realistic CPT/SPT profiles conditioned on geology, climate, or historical events (Goodfellow *et al*., 2014; Ho *et al*., 2020; Guan *et al*., 2021) [47, 60, 52]. Future applications include generating ensembles of possible subsurface scenarios for probabilistic risk assessment in earthquake induced landslides or rockfalls (Firoozi *et al*., 2024; Satipaldy *et al*., 2021) [37]. Large language models (LLMs) integrated with Geotech knowledge graphs could automate report generation, interpret ML outputs in natural language, and assist engineers in model selection (Brown *et al*., 2020; Wei *et al*., 2022) [13]. Blockchain and secure multi-party computation may ensure traceability and integrity of geotechnical datasets, critical for regulatory compliance in public infrastructure (Nakamoto, 2008; Goldfeder *et al*., 2017) [46]. These integrations collectively shift geotechnics toward proactive, autonomous systems capable of continuous learning and adaptation.

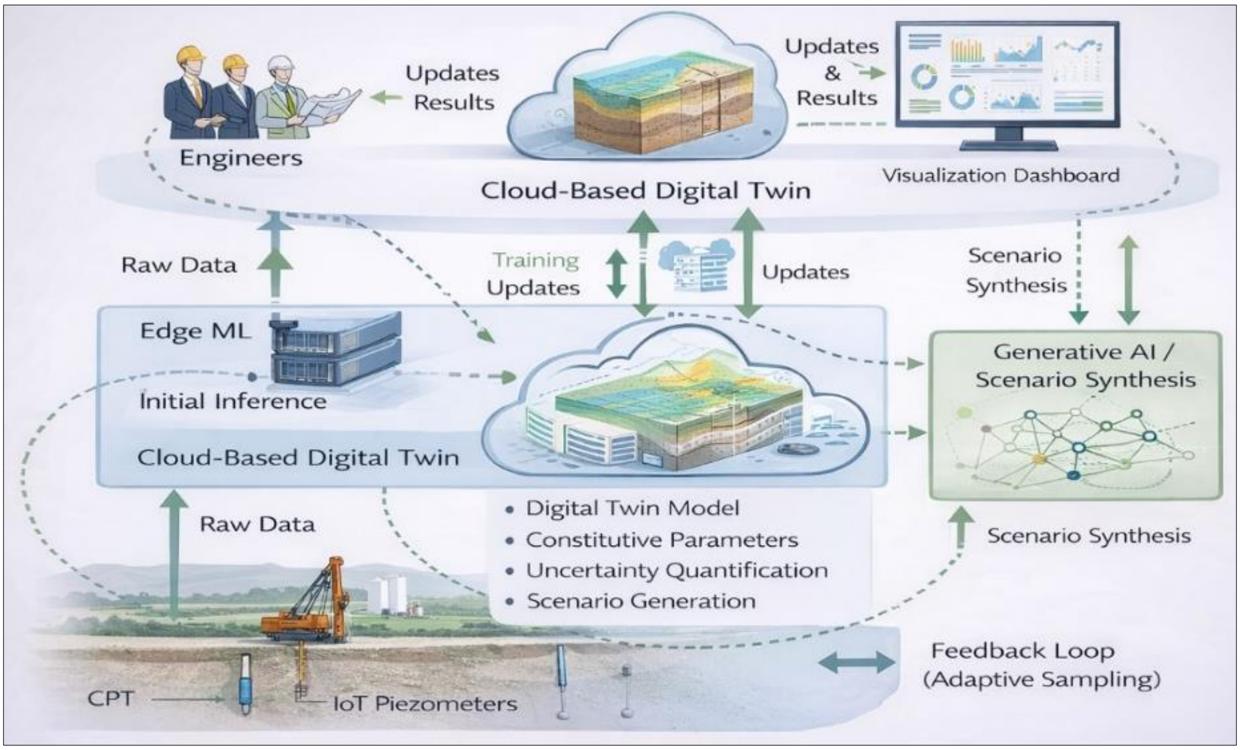

Fig 7: Conceptual architecture of a geotechnical digital twin integrated with machine learning, illustrating the flow of data from in-situ sensors (e.g., CPT and IoT piezometers) through edge-based ML for initial inference, cloud-based digital twin models for parameter updating and scenario simulation, and visualization dashboards for decision support, with feedback loops enabling adaptive site investigation and model refinement.

8.2. Advanced Paradigms

Several cutting-edge ML paradigms will dominate future developments in parameter inference and stratification.

Explainable AI (XAI) will address black-box concerns, enabling engineers to understand why models predict specific su or layer boundaries. Techniques like SHAP, LIME, and counterfactual explanations will be standard, revealing feature contributions (e.g., qc dominance in su) and detecting biases (Lundberg and Lee, 2017; Ribeiro *et al*., 2016; Mothilal *et al*., 2020). Physics informed XAI hybrids will enforce geomechanical constraints (e.g., effective stress principles) while providing interpretable saliency maps for CPT profiles (Raissi *et al*., 2019; Karniadakis *et al*., 2021) [69].

Operator learning and neural operators (e.g., Fourier Neural Operators, DeepONet) will accelerate constitutive modeling by learning mappings between function spaces, enabling fast surrogate models for complex soil behavior (Lu *et al*., 2019; Kovachki *et al*., 2021) [74]. These will replace expensive FEM runs in parameter calibration, particularly for dynamic problems like wave propagation or consolidation (Felić *et al*., 2025).

Self-supervised and foundation models pre-trained on massive unlabeled Geotech datasets will emerge, analogous to BERT or GPT in NLP. Fine-tuning on site specific CPT/SPT will yield robust inference with minimal labeled data (Devlin *et al*., 2019; Bommasani *et al*., 2021) [28]. Multimodal foundation models integrating text (reports), images (borehole photos), and tabular data (CPT logs) will provide holistic characterization (Radford *et al*., 2021).

Active learning will optimize investigation layouts: models query the most informative CPT locations to reduce uncertainty maximally (Settles, 2012; Gal *et al*., 2017) [41]. Reinforcement learning will further automate adaptive testing, balancing exploration and exploitation in real time (Sutton and Barto, 2018). These paradigms promise 50-80% reductions in required data volume and computation while enhancing reliability and interpretability.

8.3. Research Gaps

Despite progress, critical gaps remain. Standardized, open access geotechnical benchmarks beyond NGTS are lacking; curated datasets with diverse geologies, including tropical marine clays and karstic terrains, are needed for reproducible comparisons (L'Heureux and Lunne, 2020; Phoon and Zhang, 2023). Validation protocols integrating physical tests, numerical simulations, and field performance are underdeveloped (Marzouk *et al*., 2024).

Physics informed ML hybrids require better incorporation of constitutive laws (hypo plasticity, Cam Clay) and scale bridging from micro (particle level) to macro (site level) (Yuan *et al*., 2025; Karniadakis *et al*., 2021) [69]. Uncertainty propagation from inference to design (e.g., reliability indices in slopes or foundations) remains underexplored (Ching *et al*., 2022).

Interdisciplinary collaboration between geotechnical engineers, data scientists, and ethicists is insufficient; domain-specific XAI tailored to engineering judgment is needed (Rudin, 2019). Regulatory frameworks for ML in codes (e.g., Eurocode 7 updates) and certification processes are absent (EN 1997-1, 2004; Phoon, 2018) [33]. Long term performance monitoring of ML informed designs is scarce, limiting evidence of real-world benefits (Yousefpour *et al*., 2024).

8.4. Potential Impacts

The adoption of these future directions will profoundly impact geotechnical engineering and society. Enhanced prediction accuracy and uncertainty quantification will reduce over-conservatism, lowering construction costs by 10-30% while improving safety in high-risk areas (e.g., liquefaction in reclaimed lands; Bo *et al*., 2012; Firoozi *et al*., 2024) [10, 37]. Real time digital twins and adaptive sensing will enable proactive hazard mitigation, minimizing downtime in critical infrastructure like bridges and dams (Yousefpour *et al*., 2024; Nierwinski *et al*., 2025).

Sustainable development will benefit: optimized site investigations reduce drilling waste and energy use; generative AI minimizes fieldwork in environmentally sensitive zones. Climate resilience will improve through better modeling of extreme events (e.g., typhoon-induced scour or seismic rockfalls) (Satipaldy *et al*., 2021).

Broader societal gains include safer urban expansion in developing regions, accelerated infrastructure delivery, and democratization of advanced tools via open-source platforms and foundation models. By 2030, ML augmented geotechnics could become standard, akin to BIM in structural engineering, driving a paradigm shift toward intelligent, resilient earth systems.

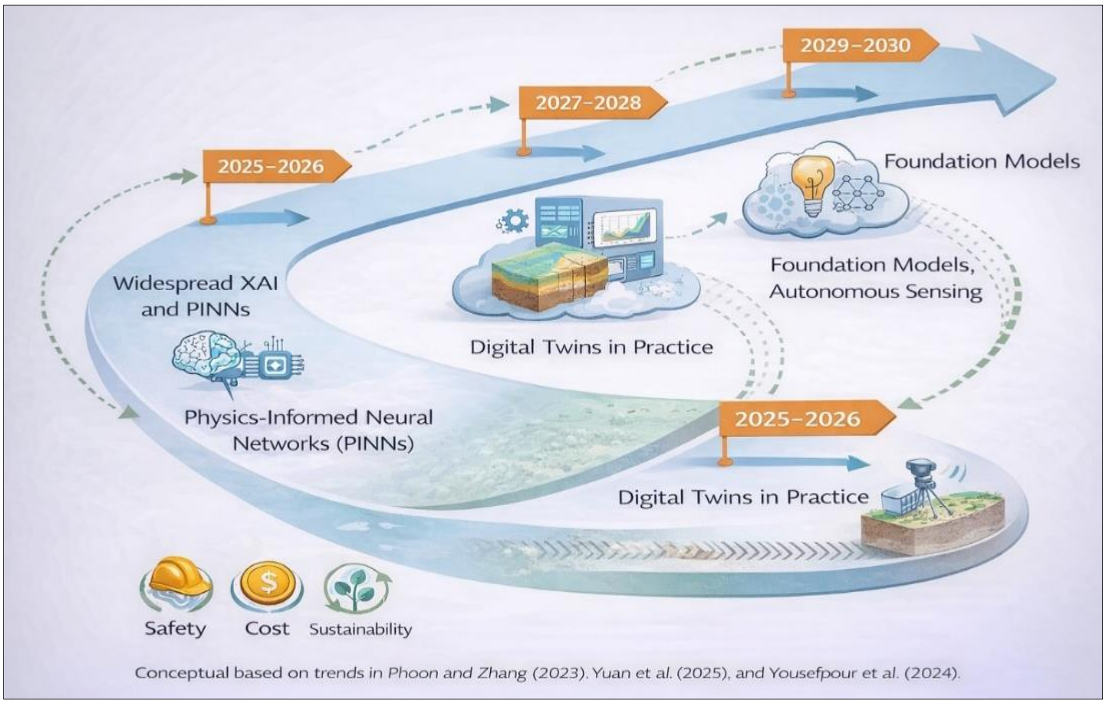

Fig 8: Roadmap to 2030 for machine learning in geotechnical site characterization, outlining the expected progression from widespread adoption of explainable AI and physics-informed neural networks, through practical implementation of digital twins, to the emergence of foundation models and autonomous sensing systems, with anticipated impacts on safety, cost efficiency, and sustainability.

9. Conclusions

Geotechnical site characterization remains one of the most critical yet challenging phases in civil engineering practice, directly influencing the safety, economy, and sustainability of infrastructure projects worldwide. Traditional empirical methods such as Robertson's Soil Behavior Type charts for CPT interpretation, Schmertmann's settlement predictions, and standard correlations for SPT derived parameters have served the profession well for decades, providing practical, experience-based tools that are simple to apply and widely accepted in design codes. However, these approaches are inherently limited by subjectivity, site specific calibration needs, scatter in correlations, and inability to fully exploit the high resolution, multivariate datasets generated by modern in-situ testing techniques. The inherent spatial variability, anisotropy, and uncertainty characteristic of natural soils particularly in complex environments such as reclaimed marine deposits, tailings impoundments, karstic terrains, or seismically active zones often lead to conservative assumptions that inflate costs or, conversely, underestimated risks that compromise performance (Phoon and Kulhawy, 1999); (Lunne *et al*., 1997); (Mayne, 2007); (Robertson, 2010).

This comprehensive review has demonstrated that advanced data-driven machine learning frameworks offer a powerful paradigm shift, enabling more accurate, efficient, and uncertainty aware soil stratification and parameter inference. From shallow supervised models (ANNs, SVR, k-NN) to deep architectures (CNNs, LSTMs), ensembles (RF, XGBoost), probabilistic methods (GPR, BNNs), and physics informed hybrids (PINNs, GDF-ML), ML approaches consistently outperform traditional empirical techniques across a wide range of metrics. Benchmarks from diverse case studies including Norwegian GeoTest Sites (NGTS), Brazilian iron tailings dams, Salzburg test sites, and urban reclamation projects show improvements in classification accuracy (F1-scores rising from 0.800-0.85 to >0.95), regression performance (R^2 from 0.65-0.80 to 0.93-0.98), and error reduction (RMSE/MAE decreases of 20-50%) for key parameters such as undrained shear strength (su), small-strain shear modulus (G0 or Vs), saturated unit weight (γsat), friction angle (φ'), and compressibility indices (Cc) (Felić *et al*., 2025) [35]; (Nierwinski *et al*., 2025); (Marzouk *et al*., 2024); (Xie *et al*., 2022); (L'Heureux and Lunne, 2020); (Yousefpour *et al*., 2024).

The reviewed frameworks excel in handling the high-dimensional, noisy, and heterogeneous nature of CPT and SPT data. Shallow learning provides interpretable baselines and performs well with modest datasets, while deep learning captures sequential and multiscale patterns in depth profiles, making it particularly effective for delineating subtle layer transitions in varved or depositional soils. Ensemble and hybrid methods enhance robustness, reduce variance, and incorporate spatial autocorrelation through innovations like geotechnical distance fields (GDFs), yielding superior continuity and generalization across sites. Probabilistic and uncertainty-aware approaches deliver credible intervals and variance estimates essential for reliability-based design, moving geotechnics closer to risk-informed decision making. Case studies further illustrate practical value: clustering algorithms reveal depositional evolution in tailings over decades, supervised regression generalizes parameters across Norwegian soft clays, and multimodal fusion improves inference in bridge scour and urban settings (Nierwinski *et al*., 2025); (Felić *et al*., 2025) [35]; (Marzouk *et al*., 2024); (Satipaldy *et al*., 2021).

Nevertheless, significant challenges persist. Data scarcity, quality issues (noise, imbalance, bias), and lack of standardized open-access benchmarks continue to limit model generalizability. Black box behavior, overfitting risks, extrapolation difficulties, and high computational demands hinder trust and real time field adoption. Interpretability gaps, regulatory lag, and integration barriers with legacy tools (e.g., PLAXIS, GeoStudio) further slow translation from research to routine practice. Ethical considerations algorithmic bias, privacy in collaborative learning, and over-reliance on automation must be proactively managed to ensure equitable and responsible deployment (Phoon and Zhang, 2023); (Rudin, 2019); (Mehrabi *et al*., 2019).

Best practices outlined in this review rigorous preprocessing, cross validation, physics informed constraints, XAI tools (SHAP, LIME), sensitivity analysis, and validation against physical tests provide a clear pathway forward. Future directions hold immense promise: digital twins integrated with IoT sensors and real time ML will enable continuous, adaptive characterization; generative AI and foundation models will alleviate data limitations through realistic synthetic profiles; operator learning and neural operators will accelerate constitutive modeling; explainable and physics informed hybrids will restore engineer trust; and active/reinforcement learning will optimize investigation layouts. These advancements, when combined with interdisciplinary collaboration and updated regulatory frameworks, could reduce over conservatism in design by 15-30%, enhance climate resilience, and support sustainable infrastructure development in an era of rapid urbanization and environmental change (Tao *et al*., 2018); (Qi *et al*., 2020); (Karniadakis *et al*., 2021) [69]; (Yuan *et al*., 2025); (Lu *et al*., 2019).

In conclusion, the integration of advanced machine learning frameworks into geotechnical site characterization marks a fundamental evolution from empirical heuristics to intelligent, data centric engineering. By synthesizing the strengths of traditional methods with the predictive power of ML, practitioners can achieve higher fidelity subsurface models, reduced uncertainties, and more resilient outcomes. While challenges remain, the trajectory is clear: continued investment in data infrastructure, interpretability, and domain-informed innovation will position ML as an indispensable tool in modern geotechnical practice. The field is poised for a new era of precision, efficiency, and safety, ultimately contributing to safer infrastructure and more sustainable development in an increasingly uncertain world.

10. References

1. Abdi H, Williams LJ. Principal component analysis. Wiley Interdiscip Rev Comput Stat. 2010;2(4):433-59.
2. Albawi S, Mohammed TA, Al-Zawi S. Understanding of a convolutional neural network. In: Proceedings of 2017 International Conference on Engineering and Technology (ICET); 2017. p. 1-6.
3. Altman NS. An introduction to kernel and nearest-neighbor nonparametric regression. Am Stat. 1992;46(3):175-85.
4. Andrus RD, Stokoe KH. Liquefaction resistance of soils from shear-wave velocity. J Geotech Geoenviron Eng. 2004;130(10):1015-25.
5. Arrieta AB, Díaz-Rodríguez N, Del Ser J, Bennetot A, Tabik S, Barbado A, *et al*. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Inf Fusion. 2020;58:82-115.
6. Barocas S, Hardt M, Narayanan A. Fairness and machine learning. fairmlbook.org; 2019.

7. Bergstra J, Bengio Y. Random search for hyper-parameter optimization. J Mach Learn Res. 2012;13:281-305.
8. Blundell C, Cornebise J, Kavukcuoglu K, Wierstra D. Weight uncertainty in neural networks. In: Proceedings of the 32nd International Conference on Machine Learning; 2015. p. 1613-22.
9. Bo MW, Arulrajah A, Horpibulsuk S, Leong M. Densification of granular soil by dynamic methods. Geotech Eng J SEAGS AGSSEA. 2015;46(4):1-12.
10. Bo MW, Arulrajah A, Horpibulsuk S, Leong M, Disfani MM. Densification of land reclamation sands by earth pressure shield tunnelling machine. Geotech Eng. 2012;43(2):1-10.
11. Boulanger RW, Idriss IM. CPT and SPT based liquefaction triggering procedures. Earthquake Spectra. 2014;30(3):1087-109.
12. Bray JD, Cubrinovski M, Zupan J, Taylor M. Liquefaction effects on buildings in the central business district of Christchurch. Bull Earthquake Eng. 2014;12(3):1167-92.
13. Brown TB, Mann B, Ryder N, Subbiah M, Kaplan J, Dhariwal P, *et al*. Language models are few-shot learners. Adv Neural Inf Process Syst. 2020;33:1877-901.
14. Chala AT, Ray RP. Machine learning techniques for soil characterization using cone penetration test data. Appl Sci. 2023;13(14):8286.
15. Chandola V, Banerjee A, Kumar V. Anomaly detection: a survey. ACM Comput Surv. 2009;41(3):15:1-15:58.
16. Chapelle O, Schölkopf B, Zien A. Semi-supervised learning. Cambridge (MA): MIT Press; 2006.
17. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic minority over-sampling technique. J Artif Intell Res. 2002;16:321-57.
18. Chen T, Guestrin C. XGBoost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2016. p. 785-94.
19. Ching J, Phoon KK, Chen YC. Reduced uncertainty of ground response using shear wave velocity measurements in the Taipei Basin. Can Geotech J. 2017;54(3):305-20.
20. Ching J, Phoon KK, Lee WT. Uncertainty quantification of geotechnical site boundaries using sparse CPT data. Georisk. 2021;15(1):1-19.
21. Ching J, Phoon KK, Sung SP. Worst case scale of fluctuation in basal loading of a circular foundation due to limited test data. Georisk. 2022;16(2):209-28.
22. Clayton CRI, Matthews MC, Simons NE. Site investigation. 2nd ed. Oxford: Blackwell Science; 1995.
23. Comaniciu D, Meer P. Mean shift: a robust approach toward feature space analysis. IEEE Trans Pattern Anal Mach Intell. 2002;24(5):603-19.
24. Cortes C, Vapnik V. Support-vector networks. Mach Learn. 1995;20(3):273-97.
25. Creswell A, White T, Dumoulin V, Arulkumaran K, Sengupta B, Bharath AA. Generative adversarial networks: an overview. arXiv preprint arXiv:1710.07035; 2018.
26. Cubrinovski M, Bray JD, Taylor M, Giorgini S, Bradley B, Wotherspoon L, *et al*. Soil liquefaction effects in the central business district during the February 2011 Christchurch earthquake. Bull N Z Soc Earthquake Eng. 2011;44(4):205-26.
27. Dean J, Ghemawat S. MapReduce: simplified data processing on large clusters. Commun ACM. 2008;51(1):107-13.
28. Devlin J, Chang MW, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for natural language understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers); 2019. p. 4171-86.
29. Dietterich TG. Ensemble methods in machine learning. In: Multiple classifier systems. Lecture Notes in Computer Science. 2000;1857:1-15.
30. Domingos P. The role of Occam's razor in knowledge discovery. Data Min Knowl Discov. 1999;3(4):409-25.
31. Drucker H, Burges CJC, Kaufman L, Smola A, Vapnik V. Support vector regression machines. Adv Neural Inf Process Syst. 1997;9.
32. Duncan JM, Wright SG, Brandon TL. Soil strength and slope stability. 2nd ed. Hoboken (NJ): John Wiley & Sons; 2014.
33. EN 1997-1. Eurocode 7: Geotechnical design - Part 1: General rules. Brussels: European Committee for Standardization; 2004.
34. Ester M, Kriegel HP, Sander J, Xu X. A density-based algorithm for discovering clusters in large spatial databases with noise. In: Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining; 1996. p. 226-31.
35. Felić H, Marzouk I, Tschuchnigg F. A data-driven approach for soil parameter determination using supervised machine learning. In: Proceedings of the 9th International Symposium on Geotechnical Safety and Risk; 2025. p. 729-33.
36. Fenton GA, Griffiths DV. Risk assessment in geotechnical engineering. Hoboken (NJ): John Wiley & Sons; 2008.
37. Firoozi AA, Firoozi AA, Atti K, Rashid MS. Integrated geotechnical-modeling and real-time analysis for predicting earthquake-induced landslides and rockfalls. Trends Ecol Indoor Environ Eng. 2024;2(3):1-19.
38. Fix E, Hodges JL. Discriminatory analysis-nonparametric discrimination: consistency properties. USA Air Force School of Aviation Medicine Report 1; 1951.
39. Friedman JH. Greedy function approximation: a gradient boosting machine. Ann Stat. 2001;29(5):1189-232.
40. Gal Y, Ghahramani Z. Dropout as a Bayesian approximation: representing model uncertainty in deep learning. In: Proceedings of the 33rd International Conference on Machine Learning; 2016. p. 1050-9.

41. Gal Y, Islam R, Ghahramani Z. Deep Bayesian active learning with image data. In: Proceedings of the 34th International Conference on Machine Learning; 2017. p. 1183-92.
42. Ganin Y, Ustinova E, Ajakan H, Germain P, Larochelle H, Marchand M, *et al*. Domain-adversarial training of neural networks. J Mach Learn Res. 2016;17(209):1-35.
43. Ghahramani Z. Probabilistic machine learning and artificial intelligence. Nature. 2015;521(7553):452-9.
44. Goh ATC. Seismic liquefaction potential assessed by neural networks. J Geotech Eng. 1994;120(9):1467-80.
45. Goldberg DE. Genetic algorithms in search, optimization, and machine learning. Reading (MA): Addison-Wesley; 1989.
46. Goldfeder S, Kaliski JA, Liskov B. Better the devil you know: exposing the data sharing practices of smartphone apps. In: Proceedings of the IEEE Symposium on Security and Privacy; 2017. p. 1-17.
47. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, *et al*. Generative adversarial nets. Adv Neural Inf Process Syst. 2014;27.
48. Goodfellow I, Bengio Y, Courville A. Deep learning. Cambridge (MA): MIT Press; 2016.
49. Graves A, Mohamed A, Hinton G. Speech recognition with deep recurrent neural networks. arXiv preprint arXiv:1303.5778; 2013.
50. Greff K, Srivastava RK, Koutník J, Steunebrink BR, Schmidhuber J. LSTM: a search space odyssey. IEEE Trans Neural Netw Learn Syst. 2017;28(10):2222-32.
51. Guo C, Pleiss G, Goldstein T, Weinberger KQ. On calibration of modern neural networks. In: Proceedings of the ICML Workshop on Reliable Machine Learning in the Wild; 2017.
52. Guan Z, Kukreja N, Singh V, Chen Y, Wang Y. GeoGAN: a conditional GAN with reconstruction and style loss to generate standard layer-specific maps of soil properties. Comput Geotech. 2021;130:103932.
53. Hastie T, Tibshirani R, Friedman J. The elements of statistical learning: data mining, inference, and prediction. 2nd ed. New York: Springer; 2009.
54. Hatanaka M, Uchida A. Empirical correlation between penetration resistance and internal friction angle of sandy soils. Soils Found. 1996;36(1):1-9.
55. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2016. p. 770-8.
56. He H, Garcia EA. Learning from imbalanced data. IEEE Trans Knowl Data Eng. 2009;21(9):1263-84.
57. Hensman J, Fusi N, Lawrence ND. Gaussian processes for big data. In: Proceedings of the 29th Conference on Uncertainty in Artificial Intelligence; 2013. p. 282-90.
58. Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. Science. 2006;313(5786):504-7.
59. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput. 1997;9(8):1735-80.
60. Ho J, Jain A, Abbeel P. Denoising diffusion probabilistic models. Adv Neural Inf Process Syst. 2020;33:6840-51.
61. Holland JH. Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence. Cambridge (MA): MIT Press; 1992.
62. Hu R, Li Y, Wang Y. Efficient estimation of probability of geological success by machine learning techniques. Bull Eng Geol Environ. 2020;79(3):1413-29.
63. Hu R, Wang Y, Zhang P, Ching J. Fully probabilistic machine learning for geotechnical site characterization. Can Geotech J. 2020;57(11):1721-36.
64. Huang J, Griffiths DV, Fenton GA. Probabilistic analysis of a discrete element modelling of the runout behavior of landslides. Georisk. 2023;17(1):1-15.
65. Idriss IM, Boulanger RW. Soil liquefaction during earthquakes. Oakland (CA): Earthquake Engineering Research Institute; 2008.
66. Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. In: Proceedings of the 32nd International Conference on Machine Learning; 2015. p. 448-56.
67. James G, Witten D, Hastie T, Tibshirani R. An introduction to statistical learning with applications in R. New York: Springer; 2013.
68. Jolliffe IT. Principal component analysis. 2nd ed. New York: Springer; 2002.
69. Karniadakis GE, Kevrekidis IG, Lu L, Perdikaris P, Wang S, Yang L. Physics-informed machine learning. Nat Rev Phys. 2021;3(6):422-40.
70. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, *et al*. LightGBM: a highly efficient gradient boosting decision tree. Adv Neural Inf Process Syst. 2017;30.
71. Kennedy J, Eberhart R. Particle swarm optimization. In: Proceedings of ICNN'95 - International Conference on Neural Networks; 1995. p. 1942-8.
72. Kingma DP, Welling M. Auto-encoding variational Bayes. arXiv preprint arXiv:1312.6114; 2014.
73. Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: Proceedings of the 14th International Joint Conference on Artificial Intelligence; 1995. p. 1137-43.
74. Kovachki N, Li Z, Liu B, Azizzadenesheli K, Bhattacharya K, Stuart A, *et al*. Neural operator: learning maps between function spaces. J Mach Learn Res. 2021;24(89):1-97.
75. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Adv Neural Inf Process Syst. 2012;25.
76. Kuhn M, Johnson K. Applied predictive modeling. New York: Springer; 2013.
77. Lacasse S, Nadim F, Hoeg K. Modelling uncertainties in landslide slope stability analysis. In: Proceedings of the TC304 ISSMGE Symposium; 2017. p. 1-10.
78. Lakshminarayanan B, Pritzel A, Blundell C. Simple and scalable predictive uncertainty estimation using deep ensembles. Adv Neural Inf Process Syst. 2017;30.
79. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521(7553):436-44.
80. Leung ECW, Bo MW, Arulrajah A. Reclamation and soil improvement using dredged hydrosediment for the

proposed western reclamation in Singapore. Geotech Eng J SEAGS AGSSEA. 2018;49(1):1-12.
81. Liaw A, Wiener M. Classification and regression by randomForest. R News. 2002;2(3):18-22.
82. Little RJA, Rubin DB. Statistical analysis with missing data. 3rd ed. Hoboken (NJ): Wiley; 2019.
83. Liu FT, Ting KM, Zhou ZH. Isolation forest. In: Proceedings of the 8th IEEE International Conference on Data Mining; 2008. p. 413-22.
84. Lloret-Cabot M, Fenton GA, Hicks MA. On the estimation of scale of fluctuation in geostatistics. Geotechnique Lett. 2014;4(3):201-6.

How to Cite This Article

Rehman MZU. Advanced data-driven machine learning frameworks for enhanced soil stratification and parameter inference in geotechnical site characterization. Int J Revolut Civ Eng. 2025 Nov–Dec;1(6):1–22. doi: 10.54660/IJRCE.2026.2.1.01-22

Creative Commons (CC) License